

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
20 March 2003 (20.03.2003)

PCT

(10) International Publication Number
WO 03/023002 A2

- (51) International Patent Classification⁷: C12N US 60/323,519 (CIP)
Filed on 19 September 2001 (19.09.2001)
- (21) International Application Number: PCT/US02/28539 US 60/323,631 (CIP)
Filed on 20 September 2001 (20.09.2001)
- (22) International Filing Date: 9 September 2002 (09.09.2002) US 60/323,636 (CIP)
Filed on 20 September 2001 (20.09.2001)
US 60/324,969 (CIP)
Filed on 25 September 2001 (25.09.2001)
- (25) Filing Language: English US 60/325,091 (CIP)
Filed on 25 September 2001 (25.09.2001)
- (26) Publication Language: English US 60/324,990 (CIP)
Filed on 26 September 2001 (26.09.2001)
US 60/373,212 (CIP)
Filed on 17 April 2002 (17.04.2002)
US 10/236,177 (CIP)
Filed on 6 September 2002 (06.09.2002)
- (30) Priority Data:
60/318,120 7 September 2001 (07.09.2001) US
60/318,130 7 September 2001 (07.09.2001) US
60/318,219 7 September 2001 (07.09.2001) US
60/318,430 10 September 2001 (10.09.2001) US
60/322,781 17 September 2001 (17.09.2001) US
60/322,636 17 September 2001 (17.09.2001) US
60/322,817 17 September 2001 (17.09.2001) US
60/322,816 17 September 2001 (17.09.2001) US
60/323,519 19 September 2001 (19.09.2001) US
60/323,631 20 September 2001 (20.09.2001) US
60/323,636 20 September 2001 (20.09.2001) US
60/324,969 25 September 2001 (25.09.2001) US
60/325,091 25 September 2001 (25.09.2001) US
60/324,990 26 September 2001 (26.09.2001) US
60/373,212 17 April 2002 (17.04.2002) US
10/236,177 6 September 2002 (06.09.2002) US
- (63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 60/318,130 (CIP)
Filed on 7 September 2001 (07.09.2001)
US 60/318,219 (CIP)
Filed on 7 September 2001 (07.09.2001)
US 60/310,430 (CIP)
Filed on 10 September 2001 (10.09.2001)
US 60/322,781 (CIP)
Filed on 17 September 2001 (17.09.2001)
US 60/322,636 (CIP)
Filed on 17 September 2001 (17.09.2001)
US 60/322,817 (CIP)
Filed on 17 September 2001 (17.09.2001)
US 60/322,816 (CIP)
Filed on 17 September 2001 (17.09.2001)
- (71) Applicant (for all designated States except US): CURAGEN CORPORATION [US/US]; 555 Long Wharf Drive, 11th floor, New Haven, CT 06511 (US).
- (72) Inventors; and
- (75) Inventors/Applicants (for US only): SPYTEK, Kimberly, A. [US/US]; 28 Court Street, number 1, New Haven, CT 06511 (US). PATTURAJAN, Meera [IN/US]; 45 Harrison Avenue, Apartment 1C, Branford, CT 06405 (US). GORMAN, Linda [US/US]; 329 Monticello Drive, Branford, CT 06405 (US). LI, Li [CN/US]; 56 Jerimoth Drive, Branford, CT 06405 (US). ANDERSON, David, W. [US/US]; 85 Montoya Drive, Branford, CT 06405 (US). ZHONG, Mei [CA/US]; 45 Harrison Avenue, Apartment 1B, Branford, CT 06405 (US). GERLACH, Valerie, L. [US/US]; 18 Rock Pasture Road, Branford, CT 06405 (US). VERNET, Corine, A., M. [FR/US]; 1739 Foxon Road, Apartment L6, Branford, CT 06471 (US). ELLERMAN, Karen [US/US]; 87 Montoya Drive, Branford, CT 06405 (US). BERGHS, Constance [NL/US]; 459 Orange Street, New Haven, CT 06511 (US). ROTHENBERG, Mark, E. [US/US]; 2 Allen Road, Clinton, CT 06413 (US). GUO, Xiaojia [CN/US]; 713 Robert Frost Drive, Branford, CT 06405 (US). SHIMKETS, Richard, A. [US/US]; 5 Indian Meadows Drive, Guilford, CT 06437 (US). LEACH, Martin, D. [GB/US]; 131 Beechwood Drive, Madison, CT 06433 (US). CATTERTON, Elina

[Continued on next page]

(54) Title: NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

(57) Abstract: Disclosed herein are nucleic acid sequences that encode novel polypeptides. Also disclosed are polypeptides encoded by these nucleic acid sequences, and antibodies that immunospecifically bind to the polypeptide, as well as derivatives, variants, mutants, or fragments of the novel polypeptide, polynucleotide, or antibody specific to the polypeptide. Vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using same are also included. The invention further discloses therapeutic, diagnostic and research methods for diagnosis, treatment, and prevention of disorders involving any one of these novel human nucleic acids and proteins.

BEST AVAILABLE COPY

WO 03/023002 A2



[FI/US]; 584 Boston Post Road, Madison, CT 06443 (US). **KEKUDA, Ramesh** [IN/US]; 71 Aiken Street, Unit R3, Norwalk, CT 06851 (US). **Ji, Weizhen** [CN/US]; 3 Business Park Drive, Room 101, Branford, CT 06405 (US). **MILLER, Charles, E.** [US/US]; 98 Saddle Hill Drive, Guilford, CT 06437 (US). **RIEGER, Daniel, K.** [DE/US]; 10A McKinnel Court, Branford, CT 06405 (US). **TAUPIER, Raymond, J., Jr.** [US/US]; 34 Pardee Place Extension, East Haven, CT 06512 (US). **SHENOY, Suresh, G.** [IN/US]; 15 Millwood Drive, Branford, CT 06405 (US). **LIU, Xiaohong** [US/US]; 360 Neponset Street, Apartment 509, Canton, MA 02021 (US). **PADIGARU, Muralidhara** [IN/US]; 71 Hampton Park, Branford, CT 06405 (US). **ALSOBROOK, John, P., II** [US/US]; 60 Lake Drive, Madison, CT 06443 (US). **LEPLEY, Denise, M.** [US/US]; 51 Church Street, Branford, CT 06405 (US). **EDINGER, Shlomit, R.** [US/US]; 766 Edgewood Avenue, New Haven, CT 06515 (US). **BURGESS, Catherine, E.** [US/US]; 90 Carriage Hill Drive, Wethersfield, CT 06109 (US).

(74) **Agent: ELRIFI, Ivor, R.**; Mintz, Levin, Cohn, Ferris, Glovsky, and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

(81) **Designated States (national):** AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.

(84) **Designated States (regional):** ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

NOVEL HUMAN PROTEINS, POLYNUCLEOTIDES ENCODING THEM AND METHODS OF USING THE SAME

FIELD OF THE INVENTION

5 The present invention relates to nucleic acids encoding proteins that are new members of the following protein families: nuclear protein-like proteins, transforming acidic coiled-coil-containing protein-like proteins, thyroid hormone receptor interactor 6-like proteins, uroporphyrinogen-III synthase-like proteins, intracellular-like proteins, LIM domain transcription factor-like proteins, voltage-dependent-calcium channel-like proteins, 10 dihydropyridine-sensitive 1-type-calcium channel-like proteins, beta-3-subunit-like proteins, nucleoporin-like proteins, BHLH protein DEC2-like proteins, keratin 18-like proteins, intracellular protein-like proteins, intracellular protein Tubby-like proteins, symplekin-like proteins, telethonin-like proteins, forkhead protein O3A-like proteins, cytochrome C-like proteins, troponin t-like proteins, XIN-like proteins, prostatic binding 15 protein-like proteins, cytoplasmic protein like homo sapiens-like proteins, zinc-finger protein HZFI-like proteins, B4-2-like proteins, Maternal effect protein staufen-like proteins, desmin like homo sapiens-like proteins, hypothetical protein-like proteins, tropomyosin alpha chain-like proteins, hermansky-pudlak syndrome-like proteins, NOT2P-like proteins, human selenium-binding-like proteins, EH domain-binding mitotic 20 phosphoprotein-like proteins, hypothetical intracellular-like proteins, MHC class I region proline rich protein-like proteins, nebullin-like proteins, golgi matrix protein GM130-like proteins, microspherule protein 1-like proteins, AK016419 mus musculus adult male testis cDNA-like proteins, utrophin (dystrophin-related protein 1)-like proteins, TPR domain-like proteins, LRR domain containing like homo sapiens-like proteins, G-rich sequence factor- 25 1-like proteins, cytoplasmic protein-like proteins, meningioma-expressed antigen 6/11 (MEA6) (MEA11)-like proteins, ancient conserved domain protein 1-like proteins, CDCRL-like proteins, HPRP-like proteins, PIBF1 protein-like proteins, cytoplasmic protein-like proteins, zinc-finger/KRAB domain containing protein-like proteins, RHO- Interactin Protein 3-like proteins, cardiac-troponin I-like protein, guanine nucleotide- 30 binding protein-like proteins, benzodiazepine receptor (BZRP) like homo sapiens-like proteins, ankyrin-repeat containing protein like homo sapiens-like proteins, acyltransferase

like homo sapiens-like proteins, GTP-binding-protein SAR1A like homo sapiens-like proteins, CGI-27 like homo sapiens-like proteins, FLJ20565 like homo sapiens-like proteins, 2410014PO7R1K like homo sapiens-like proteins, multidomain presynaptic cytomatrix protein piccola like homo sapiens-like proteins, cytosolic-sorting protein PACS-1A-like-like proteins, formin 2 like homo sapiens-like proteins, novel 5' nucleotidase-like protein-like proteins, WW domain containing protein like homo sapiens-like proteins, gasdermin like homo sapiens-like proteins, Tubby super-family protein splice variant like homo sapiens-like proteins, synaptotagmin-like protein 3-A like homo sapiens-like proteins, copine I like homo sapiens-like proteins, selenoprotein X1 like homo sapiens-like proteins, hypothetical WD-repeat like homo sapiens-like proteins, cytoplasmic protein-like proteins, TNFAIP 1-like proteins, ribosomal protein L29-like proteins, paraneoplastic antigen-like proteins, GTF21RD2 like homo sapiens-like proteins, glycolipid transfer protein-like proteins, novel copine VII-like proteins, sperm membrane protein BS-63-like proteins, FIP-2-like proteins, PEX10-like proteins.

Included in the invention are polynucleotides and the polypeptides encoded by such polynucleotides, as well as vectors, host cells, antibodies and recombinant methods for producing the polypeptides and polynucleotides, as well as methods for using the same. Methods of use encompass diagnostic and prognostic assay procedures as well as methods of treating diverse pathological conditions.

BACKGROUND OF THE INVENTION

The invention generally relates to nucleic acids and polypeptides encoded therefrom. More specifically, the invention relates to nucleic acids encoding cytoplasmic, nuclear, membrane bound, and secreted polypeptides, as well as vectors, host cells, antibodies, and recombinant methods for producing these nucleic acids and polypeptides.

SUMMARY OF THE INVENTION

The present invention is based in part on nucleic acids encoding proteins that are members of the following protein families: nuclear protein-like proteins, transforming acidic coiled-coil-containing protein-like proteins, thyroid hormone receptor interactor 6-like proteins, uroporphyrinogen-III synthase-like proteins, intracellular-like proteins, LIM domain transcription factor-like proteins, voltage-dependent-calcium channel-like proteins, dihydropyridine-sensitive L-type-calcium channel-like proteins, beta-3-subunit-like proteins, nucleoporin-like proteins, BHLH protein DEC2-like proteins, keratin 18-like proteins, intracellular protein-like proteins, intracellular protein Tubby-like proteins, symplekin-like proteins, telethonin-like proteins, forkhead protein O3A-like proteins, cytochrome C-like proteins, troponin t-like proteins, XIN-like proteins, prostatic binding protein-like proteins, cytoplasmic protein like homo sapiens-like proteins, zinc-finger protein HZF1-like proteins, B4-2-like proteins, Maternal effect protein staufer-like proteins, desmin like homo sapiens-like proteins, hypothetical protein-like proteins, tropomyosin alpha chain-like proteins, hermansky-pudlak syndrome-like proteins, NOT2P-like proteins, human selenium-binding-like proteins, EH domain-binding mitotic phosphoprotein-like proteins, hypothetical intracellular-like proteins, MHC class I region proline rich protein-like proteins, nebulin-like proteins, golgi matrix protein GM130-like proteins, microspherule protein 1-like proteins, AK016419 mus musculus adult male testis cDNA-like proteins, utrophin (dystrophin-related protein 1)-like proteins, TPR domain-like proteins, LRR domain containing like homo sapiens-like proteins, G-rich sequence factor-1-like proteins, cytoplasmic protein-like proteins, meningioma-expressed antigen 6/11 (MEA6) (MEA11)-like proteins, ancient conserved domain protein 1-like proteins, CDCRL-like proteins, HPRP-like proteins, PIBF1 protein-like proteins, cytoplasmic protein-like proteins, zinc-finger/KRAB domain containing protein-like proteins, RHO-Interactin Protein 3-like proteins, cardiac-troponin I-like protein, guanine nucleotide-binding protein-like proteins,

benzodiazepine receptor (BZRP) like homo sapiens-like proteins, ankyrin-repeat containing protein like homo sapiens-like proteins, acyltransferase like homo sapiens-like proteins, GTP-binding-protein SAR1A like homo sapiens-like proteins, CGI-27 like homo sapiens-like proteins, FLJ20565 like homo sapiens-like proteins, 2410014PO7R1K like homo sapiens-like proteins, multidomain presynaptic cytomatrix protein piccola like homo sapiens-like proteins, cytosolic-sorting protein PACS-1A-like-like proteins, formin 2 like homo sapiens-like proteins, novel 5' nucleotidase-like protein-like proteins, WW domain containing protein like homo sapiens-like proteins, gasdermin like homo sapiens-like proteins, Tubby super-family protein splice variant like homo sapiens-like proteins, synaptotagmin-like protein 3-A like homo sapiens-like proteins, copine I like homo sapiens-like proteins, selenoprotein XI like homo sapiens-like proteins, hypothetical WD-repeat like homo sapiens-like proteins, cytoplasmic protein-like proteins, TNFAIP 1-like proteins, ribosomal protein L29-like proteins, paraneoplastic antigen-like proteins, GTF21RD2 like homo sapiens-like proteins, glycolipid transfer protein-like proteins, novel copine VII-like proteins, sperm membrane protein BS-63-like proteins, FIP-2-like proteins, PEX10-like proteins. The novel polynucleotides and polypeptides are referred to herein as NOV1a, NOV2a, NOV2b, NOV3a, NOV3b, NOV4a, NOV4b, NOV5a, NOV5b, NOV6a, NOV6b, NOV7a, NOV7b, NOV7c, NOV7d, NOV7e, NOV8a, NOV9a, NOV9b, NOV10a, NOV10b, NOV11a, NOV12a, NOV12b, NOV13a, NOV14a, NOV15a, NOV15b, NOV16a, NOV17a, NOV18a, NOV18b, NOV18c, NOV19a, NOV19b, NOV20a, NOV20b, NOV20c, NOV20d, NOV20e, NOV20f, NOV20g, NOV21a, NOV21b, NOV22a, NOV23a, NOV23b, NOV24a, NOV25a, NOV25b, NOV26a, NOV26b, NOV26c, NOV27a, NOV27b, NOV28a, NOV28b, NOV28c, NOV28d, NOV28e, NOV28f, NOV29a, NOV29b, NOV30a, NOV31a, NOV32a, NOV32b, NOV33a, NOV34a, NOV35a, NOV35b, NOV35c, NOV36a, NOV36b, NOV37a, NOV37b, NOV37c, NOV38a, NOV39a, NOV40a, NOV41a, NOV42a, NOV42b, NOV43a, NOV44a, NOV44b, NOV44c, NOV45a, NOV46a, NOV47a, NOV48a, NOV48b, NOV49a, NOV49b, NOV50a, NOV50b, NOV50c, NOV51a, NOV52a, NOV52b, NOV53a, NOV54a, NOV54b, NOV55a, NOV55b, NOV55c, NOV55d, NOV55e, NOV56a, NOV57a, NOV58a, NOV59a, NOV60a, NOV61a, NOV62a, NOV62b, NOV63a, NOV63b, NOV64a, NOV65a, NOV66a, NOV66b, NOV67a, NOV68a, NOV69a, NOV70a, NOV71a, NOV72a, NOV72b, NOV72c, NOV73a, NOV73b, NOV74a, NOV75a, NOV76a, NOV77a, NOV78a, NOV79a, NOV80a, NOV81a, NOV81b, NOV82a, NOV82b, NOV82c, NOV83a, NOV83b, NOV84a, NOV84b, NOV84c, NOV85a, NOV85b, NOV86a, NOV86b, NOV87a, NOV87b, NOV87c, NOV87d,

NOV87e, NOV88a, NOV88b, NOV89a, NOV89b, NOV90a, NOV90b, NOV91a, NOV91b, NOV91c, NOV91d, NOV92a, NOV92b, NOV92c and NOV92d. These nucleic acids and polypeptides, as well as derivatives, homologs, analogs and fragments thereof, will hereinafter be collectively designated as "NOVX" nucleic acid or polypeptide sequences.

5 In one aspect, the invention provides an isolated NOVX nucleic acid disclosed in SEQ ID NO:2n-1, wherein n is an integer between 1 and 172. In some embodiments, the NOVX nucleic acid molecule will hybridize under stringent conditions to a nucleic acid sequence complementary to a nucleic acid molecule that includes a protein-coding sequence of a NOVX nucleic acid sequence. The invention also includes an isolated nucleic acid that
10 encodes a NOVX polypeptide, or a fragment, homolog, analog or derivative thereof. For example, the nucleic acid can encode a polypeptide at least 80% identical to a polypeptide comprising the amino acid sequences of SEQ ID NO:2n, wherein n is an integer between 1 and 172. The nucleic acid can be, for example, a genomic DNA fragment or a cDNA molecule that includes the nucleic acid sequence of any of SEQ ID NO:2n-1, wherein n is an
15 integer between 1 and 172. Also included in the invention is an oligonucleotide, *e.g.*, an oligonucleotide which includes at least 6 contiguous nucleotides of a NOVX nucleic acid (*e.g.*, SEQ ID NO:2n-1, wherein n is an integer between 1 and 172) or a complement of said oligonucleotide.

The invention also encompasses isolated NOVX polypeptides (SEQ ID NO:2n, wherein n is an integer between 1 and 172). In certain embodiments, the NOVX
20 polypeptides include an amino acid sequence that is substantially identical to the amino acid sequence of a human NOVX polypeptide.

The invention also features antibodies that immunoselectively bind to NOVX polypeptides, or fragments, homologs, analogs or derivatives thereof.

25 In another aspect, the invention includes pharmaceutical compositions that include therapeutically- or prophylactically-effective amounts of a therapeutic and a pharmaceutically-acceptable carrier. The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or an antibody specific for a NOVX polypeptide. In a further aspect, the invention includes, in one or more containers, a therapeutically- or prophylactically-effective
30 amount of this pharmaceutical composition.

In a further aspect, the invention includes a method of producing a polypeptide by culturing a cell that includes a NOVX nucleic acid, under conditions allowing for expression

of the NOVX polypeptide encoded by the DNA. If desired, the NOVX polypeptide can then be recovered.

In another aspect, the invention includes a method of detecting the presence of a NOVX polypeptide in a sample. In the method, a sample is contacted with a compound that
5 selectively binds to the polypeptide under conditions allowing for formation of a complex between the polypeptide and the compound. The complex is detected, if present, thereby identifying the NOVX polypeptide within the sample.

The invention also includes methods to identify specific cell or tissue types based on their expression of a NOVX.

10 Also included in the invention is a method of detecting the presence of a NOVX nucleic acid molecule in a sample by contacting the sample with a NOVX nucleic acid probe or primer, and detecting whether the nucleic acid probe or primer bound to a NOVX nucleic acid molecule in the sample.

In a further aspect, the invention provides a method for modulating the activity of a
15 NOVX polypeptide by contacting a cell sample that includes the NOVX polypeptide with a compound that binds to the NOVX polypeptide in an amount sufficient to modulate the activity of said polypeptide. The compound can be, *e.g.*, a small molecule, such as a nucleic acid, peptide, polypeptide, peptidomimetic, carbohydrate, lipid or other organic (carbon containing) or inorganic molecule, as further described herein.

20 In another embodiment, the invention involves a method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of a polypeptide with an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172, the method including providing a cell expressing the polypeptide of the
25 invention and having a property or function ascribable to the polypeptide; contacting the cell with a composition comprising a candidate substance; and determining whether the substance alters the property or function ascribable to the polypeptide; whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition devoid of the substance, the substance is identified as a potential therapeutic
30 agent.

Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing disorders or syndromes including, *e.g.*, adrenoleukodystrophy, congenital adrenal hyperplasia, hemophilia, hypercoagulation,

hypogonadism, idiopathic thrombocytopenic purpura, autoimmune disease, inflammatory bowel disease (IBD), rheumatoid arthritis, osteoarthritis, psoriasis, allergies, asthma, immunodeficiencies, Von Hippel-Lindau (VHL) syndrome, Alzheimer's disease, stroke, tuberous sclerosis, hypercalcemia, Parkinson's disease, Huntington's disease, cerebral palsy, epilepsy, Lesch-Nyhan syndrome, multiple sclerosis, schizophrenia, depression, ataxia-telangiectasia, leukodystrophies, behavioral disorders, addiction, anxiety, pain, obesity, diabetes, renal artery stenosis, interstitial nephritis, glomerulonephritis, polycystic kidney disease, systemic lupus erythematosus, renal tubular acidosis, IgA nephropathy, emphysema, scleroderma, adult respiratory distress syndrome (ARDS), lymphedema, graft versus host disease (GVHD), pancreatitis, ulcers, anemia, ataxia-telangiectasia, cancer, trauma, viral infections, bacterial infections, parasitic infections; and conditions related to transplantation, neuroprotection, fertility, or regeneration (in vitro and in vivo) and/or other pathologies and disorders of the like. Also within the scope of the invention is the use of a therapeutic in the manufacture of a medicament for treating or preventing conditions including, *e.g.*, those associated with homologs of a NOVX sequence, such as those listed in Table A.

The therapeutic can be, *e.g.*, a NOVX nucleic acid, a NOVX polypeptide, or a NOVX-specific antibody, or biologically-active derivatives or fragments thereof.

For example, the compositions of the present invention will have efficacy for treatment of patients suffering from the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The polypeptides can be used as immunogens to produce antibodies specific for the invention, and as vaccines. They can also be used to screen for potential agonist and antagonist compounds. For example, a cDNA encoding NOVX may be useful in gene therapy, and NOVX may be useful when administered to a subject in need thereof.

The invention further includes a method for screening for a modulator of disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. The method includes contacting a test compound with a NOVX polypeptide and determining if the test compound binds to said NOVX polypeptide. Binding of the test compound to the NOVX polypeptide indicates the test compound is a modulator of activity, or of latency or predisposition to the aforementioned disorders or syndromes.

Also within the scope of the invention is a method for screening for a modulator of activity, or of latency or predisposition to disorders or syndromes including, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like by administering a test compound to a test animal at increased risk for the aforementioned disorders or syndromes. The test animal expresses a recombinant polypeptide encoded by a NOVX nucleic acid. Expression or activity of NOVX polypeptide is then measured in the test animal, as is expression or activity of the protein in a control animal which recombinantly-expresses NOVX polypeptide and is not at increased risk for the disorder or syndrome. Next, the expression of NOVX polypeptide in both the test animal and the control animal is compared. A change in the activity of NOVX polypeptide in the test animal relative to the control animal indicates the test compound is a modulator of latency of the disorder or syndrome.

In yet another aspect, the invention includes a method for determining the presence of or predisposition to a disease associated with altered levels of a NOVX polypeptide, a NOVX nucleic acid, or both, in a subject (*e.g.*, a human subject). The method includes measuring the amount of the NOVX polypeptide in a test sample from the subject and comparing the amount of the polypeptide in the test sample to the amount of the NOVX polypeptide present in a control sample. An alteration in the level of the NOVX polypeptide in the test sample as compared to the control sample indicates the presence of or predisposition to a disease in the subject. Preferably, the predisposition includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like. Also, the expression levels of the new polypeptides of the invention can be used in a method to screen for various cancers as well as to determine the stage of cancers.

In a further aspect, the invention includes a method of treating or preventing a pathological condition associated with a disorder in a mammal by administering to the subject a NOVX polypeptide, a NOVX nucleic acid, or a NOVX-specific antibody to a subject (*e.g.*, a human subject), in an amount sufficient to alleviate or prevent the pathological condition. In preferred embodiments, the disorder, includes, *e.g.*, the diseases and disorders disclosed above and/or other pathologies and disorders of the like.

In yet another aspect, the invention can be used in a method to identify the cellular receptors and downstream effectors of the invention by any one of a number of techniques commonly employed in the art. These include but are not limited to the two-hybrid system, affinity purification, co-precipitation with antibodies or other specific-interacting molecules.

NOVX nucleic acids and polypeptides are further useful in the generation of antibodies that bind immuno-specifically to the novel NOVX substances for use in therapeutic or diagnostic methods. These NOVX antibodies may be generated according to methods known in the art, using prediction from hydrophobicity charts, as described in the

5 "Anti-NOVX Antibodies" section below. The disclosed NOVX proteins have multiple hydrophilic regions, each of which can be used as an immunogen. These NOVX proteins can be used in assay systems for functional analysis of various human disorders, which will help in understanding of pathology of the disease and development of new drug targets for various disorders.

10 The NOVX nucleic acids and proteins identified here may be useful in potential therapeutic applications implicated in (but not limited to) various pathologies and disorders as indicated below. The potential therapeutic applications for this invention include, but are not limited to: protein therapeutic, small molecule drug target, antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), diagnostic and/or prognostic marker, gene

15 therapy (gene delivery/gene ablation), research tools, tissue regeneration *in vivo* and *in vitro* of all tissues and cell types composing (but not limited to) those defined here.

Unless otherwise defined, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the art to which this invention belongs. Although methods and materials similar or equivalent to those described herein can

20 be used in the practice or testing of the present invention, suitable methods and materials are described below. All publications, patent applications, patents, and other references mentioned herein are incorporated by reference in their entirety. In the case of conflict, the present specification, including definitions, will control. In addition, the materials, methods, and examples are illustrative only and not intended to be limiting.

25 Other features and advantages of the invention will be apparent from the following detailed description and claims.

DETAILED DESCRIPTION OF THE INVENTION

The present invention provides novel nucleotides and polypeptides encoded thereby. Included in the invention are the novel nucleic acid sequences, their encoded polypeptides,

30 antibodies, and other related compounds. The sequences are collectively referred to herein as "NOVX nucleic acids" or "NOVX polynucleotides" and the corresponding encoded polypeptides are referred to as "NOVX polypeptides" or "NOVX proteins." Unless indicated

otherwise, "NOVX" is meant to refer to any of the novel sequences disclosed herein. Table A provides a summary of the NOVX nucleic acids and their encoded polypeptides.

TABLE A. Sequences and Corresponding SEQ ID Numbers

NOVX Assignment	Internal Identification	SEQ ID NO (nucleic acid)	SEQ ID NO (amino acid)	Homology
1a	CG101036-01	1	2	Nuclear Protein
2a	CG101055-01	3	4	Transforming acidic coiled-coil-containing protein
2b	CG101055-02	5	6	
3a	CG101973-01	7	8	Thyroid Hormone Receptor Interactor 6
3b	CG101973-02	9	10	
4a	CG102244-01	11	12	Uroporphyrinogen-III Synthase
4b	CG102244-02	13	14	
5a	CG102731-01	15	16	intracellular
5b	CG102713-02	17	18	
6a	CG102975-01	19	20	LIM Domain Transcription Factor
6b	CG102975-02	21	22	
7a	CG103764-01	23	24	Voltage-dependent calcium channel, public
7b	CG103764-01	25	26	Dihydropyridine-Sensitive L-type, Calcium Channel Beta-3 Subunit-like Proteins
7c	212779035	27	28	Dihydropyridine-Sensitive L-type, Calcium Channel Beta-3 Subunit-like Proteins
7d	CG103764-02	29	30	
7e	CG103764-03	31	32	
8a	CG104944-01	33	34	Nucleoporin
9a	CG106550-01	35	36	
9b	CG106550-02	37	38	BHLH Protein DEC2
10a	CG106842-01	39	40	Keratin 18
10b	CG106842-02	41	42	
11a	CG107095-01	43	44	Intracellular
12a	CG107477-01	45	46	
12b	CG107477-02	47	48	Intracellular protein
13a	CG108707-01	49	50	Intracellular Protein Tubby
14a	CG108791-01	51	52	Symplekin
15a	CG109247-01	53	53	
15b	CG109247-02	55	56	Telethonin
16a	CG110410-01	57	58	Forkhead Protein O3A
17a	CG110882-01	59	60	Cytochrome C
18a	CG111188-01	61	62	kinectin
18b	CG111188-02	63	64	
18c	CG111188-03	65	66	
19a	CG111473-01	67	68	troponin T-like protein
19b	CG111473-02	69	70	
20a	CG111501-01	71	72	XIN

20b	CG111501-02	73	74	XIN
20c	249257832	75	76	XIN
20d	249263153	77	78	XIN
20e	249263166	79	80	XIN
20f	249263170	81	82	XIN
20g	CG111501-03	83	84	
21a	CG112595-01	85	86	Prostatic Binding Protein
21b	CG112595-02	87	88	
22a	CG112624-01	89	90	Cytoplasmic Protein like homo sapiens
23a	CG113823-01	91	92	Zinc Finger Protein HZF1
23b	CG113823-02	93	94	
24a	CG114098-01	95	96	B4-2
25a	CG114308-01	97	98	Maternal Effect Protein Staufien
25b	CG114308-02	99	100	
26a	CG114349-01	101	102	
26b	CG114349-02	103	104	desmin like homo sapiens
26c	CG114349-03	105	106	
27a	CG114503-01	107	108	hypothetical protein
27b	CG114503-02	109	110	
28a	CG114588-01	111	112	Tropomyosin Alpha Chain
28b	CG114588-02	113	114	
28c	CG114588-03	115	116	
28d	CG114588-04	117	118	
28e	CG114588-05	119	120	
28f	CG114588-06	121	122	
29a	CG114621-01	123	124	Hermansky-Pudlak Syndrome
29b	CG114621-02	125	126	
30a	CG114649-01	127	128	NOT2P
31a	CG116785-01	129	130	Human selenium-binding
32a	CG118927-01	131	132	
32b	CG118927-02	133	134	EH Domain-Binding Mitotic Phosphoprotein
33a	CG118981-01	135	136	Hypothetical Intracellular
34a	CG119385-01	137	138	MHC class I region proline rich protein
35a	CG119566-01	139	140	Nebullin
35b	CG119566-02	141	142	
35c	CG119566-03	143	144	
36a	CG120166-01	145	146	
36b	CG120166-02	147	148	Golgi matrix protein GM130
37a	CG120401-01	149	150	
37b	CG120401-02	151	152	Microspherule Protein 1
37c	CG120401-03	153	154	
38a	CG122125-01	155	156	AK016419 Mus musculus adult male testis cDNA
39a	CG122195-01	157	158	Utrophin (Dystrophin-Related Protein 1)
40a	CG122738-01	159	160	TPR domain
41a	CG123451-01	161	162	Intracellular Protein

42a	CG123660-01	163	164	Transforming Acidic Coiled-Coil-Containing Protein 2
42b	CG123660-02	165	166	
43a	CG123955-01	167	168	Zinc finger protein
44a	CG124672-01	169	170	
44b	CG124672-03	171	172	LRR Domain Containing like homo sapiens
44c	CG124672-02	173	174	LRR Domain Containing like homo sapiens
45a	CG125900-01	175	176	G-Rich Sequence Factor-1
46a	CG126510-01	177	178	Cytoplasmic protein
47a	CG127106-01	179	180	Meningioma-Expressed Antigen 6/11 (MEA6) (MEA11)
48a	CG127340-01	181	182	
48b	CG127340-02	183	184	Ancient Conserved Domain Protein 1
49a	CG128310-01	185	186	CDCRL
49b	CG128310-02	187	188	
50a	CG128369-01	189	190	HRP
50b	CG128369-02	191	192	
50c	CG128369-03	193	194	
51a	CG128420-01	195	196	PIBF1 protein
52a	CG128519-01	197	198	Cytoplasmic protein
52b	CG128519-02	199	200	
53a	CG128626-01	201	202	Zinc Finger / KRAB domain containing Protein
54a	CG128852-01	203	204	RHO-Interacting Protein 3
54b	CG128852-02	205	206	
55a	CG132650-01	207	208	
55b	CG132650-05	209	210	cardiac troponin I
55c	CG132650-02	211	212	
55d	CG132650-03	213	214	
55e	CG132650-04	215	216	
56a	CG133808-01	217	218	
57a	CG136288-01	219	220	Guanine Nucleotide-Binding Protein Gamma-7 Subunit like homo sapiens
58a	CG136933-01	221	222	2410017P07RIK like homo sapiens
59a	CG136942-01	223	224	FLJ20565 like homo sapiens
60a	CG137017-01	225	226	CGI-27 like homo sapiens
61a	CG137146-01	227	228	GTP-Binding Protein SAR1A like homo sapiens
62a	CG137566-01	229	230	Acyltransferase like homo sapiens
62b	CG137566-02	231	232	
63a	CG137707-01	233	234	Benzodiazpine receptor (BZRP) like homo sapiens
63b	CG137707-02	235	236	
64a	CG138033-01	237	238	Ankyrin-repeat containing protein like homo sapiens

65a	CG138043-01	239	240	Multidomain Presynaptic Cytomatrix Protein Piccolo like homo sapiens
66a	CG138208-01	241	242	Cytosolic Sorting Protein PACS-1A-like
66b	CG138208-02	243	244	Cytosolic Sorting Protein PACS-1A-like
67a	CG138303-01	245	246	Formin 2 like homo sapiens
68a	CG138362-01	247	248	Novel 5' nucleotidase-like Proteins
69a	CG138452-01	249	250	Novel Intracellular F-box domain containing protein-like Proteins and Nucleic Acids Encoding Same
70a	CG138781-01	251	252	Novel HTPHLP Gene Like-like Proteins and Nucleic Acids Encoding Same
71a	CG138808-01	253	254	Novel SIF AND TIAMI-Like Exchange Factor-like Proteins and Nucleic Acids Encoding Same
72a	CG139224-01	255	256	WW domain containing protein like homo sapiens
72b	CG139224-02	257	258	
72c	CG139224-03	259	260	
73a	CG140088-01	261	262	Gasdermin like homo sapiens
73b	CG140088-02	263	264	
74a	CG140170-01	265	266	Tubby Super-Family Protein Splice Variant like homo sapiens
75a	CG140179-01	267	268	Synaptotagmin-Like Protein 3-A like homo sapiens
76a	CG140392-01	269	270	Hypothetical Intracellular like homo sapiens
77a	CG140727-01	271	272	Copine I like homo sapiens
78a	CG141070-01	273	274	Selenoprotein X 1 like homo sapiens
79a	CG141395-01	275	276	Hypothetical WD-repeat like homo sapiens
80a	CG191018-01	277	278	cytoplasmic protein
81a	CG56125-01	279	280	
81b	CG56125-02	281	282	TNFAIP1
82a	CG57113-01	283	284	
82b	CG57113-03	285	286	Ribosomal Protein L29
82c	CG57113-02	287	288	
83a	CG59536-01	289	290	
83b	CG59536-02	291	292	Paraneoplastic Antigen
84a	CG59794-01	293	294	GTF2IRD2 like homo sapiens
84b	CG59794-02	295	296	GTF2IRD2 like homo sapiens
84c	CG59794-03	297	298	
85a	CG59821-01	299	300	Intracellular Protein like homo sapiens

85b	CG59821-02	301	302	Intracellular Protein like homo sapiens
86a	CG59849-01	303	304	
86b	CG49849-02	305	306	Intracellular Protein like homo sapiens
87a	CG59920-01	307	308	Glycolipid Transfer Protein
87b	CG59920-02	309	310	Glycolipid Transfer Protein
87c	277583551	311	312	Glycolipid Transfer Protein
87d	CG59920-01	313	314	Glycolipid Transfer Protein
87e	308559628	315	316	Glycolipid Transfer Protein
88a	CG59983-01	317	318	Novel Copine VII-like Proteins
88b	CG59983-02	319	320	Novel Copine VII-like Proteins
89a	CG93335-01	321	322	
89b	CG93335-02	323	324	Intracellular protein
90a	CG94377-01	325	326	
90b	CG94377-02	327	328	Sperm Membrane Protein BS-63
91a	CG97090-01	329	330	
91b	CG97090-04	331	332	FIP-2
91c	CG97090-03	333	334	FIP-2
91d	CG97090-02	335	336	
92a	CG97966-01	337	338	
92b	CG97966-03	339	340	PEX10
92c	CG97966-02	341	342	
92d	CG97966-04	343	344	

Table A indicates the homology of NOV polypeptides to known protein families.

Thus, the nucleic acids and polypeptides, antibodies and related compounds according to the invention corresponding to a NOVX as identified in column 1 of Table A will be useful in therapeutic and diagnostic applications implicated in, for example, pathologies and disorders associated with the known protein families identified in column 5 of Table A.

Pathologies, diseases, disorders and condition and the like that are associated with NOVX sequences include, but are not limited to: *e.g.*, cardiomyopathy, atherosclerosis, hypertension, congenital heart defects, aortic stenosis, atrial septal defect (ASD), atrioventricular (A-V) canal defect, ductus arteriosus, pulmonary stenosis, subaortic stenosis, ventricular septal defect (VSD), valve diseases, tuberous sclerosis, scleroderma, obesity, metabolic disturbances associated with obesity, transplantation, adrenoleukodystrophy, congenital adrenal hyperplasia, prostate cancer, pancreatic cancer, gastric cancer, colon cancer, liver cancer, renal cancer, breast cancer, ovarian cancer, prostate cancer, squamous cell carcinoma, melanoma, brain cancer, allergies, asthma, emphysema, inflammatory bowel disease, rheumatoid arthritis, osteoarthritis, lupus erythematosus, diabetes, metabolic

disorders, neoplasm; adenocarcinoma, lymphoma, uterus cancer, fertility, hemophilia, hypercoagulation, idiopathic thrombocytopenic purpura, immunodeficiencies, graft versus host disease, AIDS, bronchial asthma, Crohn's disease; multiple sclerosis, schizophrenia, depression, treatment of Albright Hereditary Osteodystrophy, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, epilepsy, Alzheimer's Disease, Parkinson's Disorder, Huntington's Disease, immune disorders, hematopoietic disorders, and the various dyslipidemias, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers, as well as conditions such as transplantation and fertility.

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

Consistent with other known members of the family of proteins, identified in column 5 of Table A, the NOVX polypeptides of the present invention show homology to, and contain domains that are characteristic of, other members of such protein families. Details of the sequence relatedness and domain analysis for each NOVX are presented in Example A.

The NOVX nucleic acids and polypeptides can also be used to screen for molecules, which inhibit or enhance NOVX activity or function. Specifically, the nucleic acids and polypeptides according to the invention may be used as targets for the identification of small molecules that modulate or inhibit diseases associated with the protein families listed in Table A.

The NOVX nucleic acids and polypeptides are also useful for detecting specific cell types. Details of the expression analysis for each NOVX are presented in Example C. Accordingly, the NOVX nucleic acids, polypeptides, antibodies and related compounds according to the invention will have diagnostic and therapeutic applications in the detection of a variety of diseases with differential expression in normal vs. diseased tissues, *e.g.* detection of a variety of cancers.

Additional utilities for NOVX nucleic acids and polypeptides according to the invention are disclosed herein.

NOVX clones

NOVX nucleic acids and their encoded polypeptides are useful in a variety of applications and contexts. The various NOVX nucleic acids and polypeptides according to the invention are useful as novel members of the protein families according to the presence of domains and sequence relatedness to previously described proteins. Additionally, NOVX nucleic acids and polypeptides can also be used to identify proteins that are members of the family to which the NOVX polypeptides belong.

The NOVX genes and their corresponding encoded proteins are useful for preventing, treating or ameliorating medical conditions, *e.g.*, by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the NOVX genes, based on the tissues in which they are most highly expressed. Uses include developing products for the diagnosis or treatment of a variety of diseases and disorders.

The NOVX nucleic acids and proteins of the invention are useful in potential diagnostic and therapeutic applications and as a research tool. These include serving as a specific or selective nucleic acid or protein diagnostic and/or prognostic marker, wherein the presence or amount of the nucleic acid or the protein are to be assessed, as well as potential therapeutic applications such as the following: (i) a protein therapeutic, (ii) a small molecule drug target, (iii) an antibody target (therapeutic, diagnostic, drug targeting/cytotoxic antibody), (iv) a nucleic acid useful in gene therapy (gene delivery/gene ablation), and (v) a composition promoting tissue regeneration *in vitro* and *in vivo* (vi) a biological defense weapon.

In one specific embodiment, the invention includes an isolated polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172, wherein any amino acid in the mature form is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) an amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer

between 1 and 172 wherein any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; and (e) a fragment of any of (a) through (d).

In another specific embodiment, the invention includes an isolated nucleic acid molecule comprising a nucleic acid sequence encoding a polypeptide comprising an amino acid sequence selected from the group consisting of: (a) a mature form of the amino acid sequence given SEQ ID NO: 2n, wherein n is an integer between 1 and 172; (b) a variant of a mature form of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172 wherein any amino acid in the mature form of the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence of the mature form are so changed; (c) the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172; (d) a variant of the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172, in which any amino acid specified in the chosen sequence is changed to a different amino acid, provided that no more than 15% of the amino acid residues in the sequence are so changed; (e) a nucleic acid fragment encoding at least a portion of a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO: 2n, wherein n is an integer between 1 and 172 or any variant of said polypeptide wherein any amino acid of the chosen sequence is changed to a different amino acid, provided that no more than 10% of the amino acid residues in the sequence are so changed; and (f) the complement of any of said nucleic acid molecules.

In yet another specific embodiment, the invention includes an isolated nucleic acid molecule, wherein said nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of: (a) the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 172; (b) a nucleotide sequence wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 172 is changed from that selected from the group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed; (c) a nucleic acid fragment of the sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 172; and (d) a nucleic acid fragment wherein one or more nucleotides in the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 172 is changed from that selected from the

group consisting of the chosen sequence to a different nucleotide provided that no more than 15% of the nucleotides are so changed.

NOVX Nucleic Acids and Polypeptides

One aspect of the invention pertains to isolated nucleic acid molecules that encode
5 NOVX polypeptides or biologically active portions thereof. Also included in the invention are nucleic acid fragments sufficient for use as hybridization probes to identify NOVX-encoding nucleic acids (*e.g.*, NOVX mRNAs) and fragments for use as PCR primers for the amplification and/or mutation of NOVX nucleic acid molecules. As used herein, the term “nucleic acid molecule” is intended to include DNA molecules (*e.g.*, cDNA or genomic
10 DNA), RNA molecules (*e.g.*, mRNA), analogs of the DNA or RNA generated using nucleotide analogs, and derivatives, fragments and homologs thereof. The nucleic acid molecule may be single-stranded or double-stranded, but preferably is comprised double-stranded DNA.

A NOVX nucleic acid can encode a mature NOVX polypeptide. As used herein, a
15 “mature” form of a polypeptide or protein disclosed in the present invention is the product of a naturally occurring polypeptide or precursor form or proprotein. The naturally occurring polypeptide, precursor or proprotein includes, by way of nonlimiting example, the full-length gene product encoded by the corresponding gene. Alternatively, it may be defined as the polypeptide, precursor or proprotein encoded by an ORF described herein. The product
20 “mature” form arises, by way of nonlimiting example, as a result of one or more naturally occurring processing steps that may take place within the cell (*e.g.*, host cell) in which the gene product arises. Examples of such processing steps leading to a “mature” form of a polypeptide or protein include the cleavage of the N-terminal methionine residue encoded by the initiation codon of an ORF, or the proteolytic cleavage of a signal peptide or leader
25 sequence. Thus a mature form arising from a precursor polypeptide or protein that has residues 1 to N, where residue 1 is the N-terminal methionine, would have residues 2 through N remaining after removal of the N-terminal methionine. Alternatively, a mature form arising from a precursor polypeptide or protein having residues 1 to N, in which an N-terminal signal sequence from residue 1 to residue M is cleaved, would have the residues
30 from residue M+1 to residue N remaining. Further as used herein, a “mature” form of a polypeptide or protein may arise from a step of post-translational modification other than a proteolytic cleavage event. Such additional processes include, by way of non-limiting

example, glycosylation, myristylation or phosphorylation. In general, a mature polypeptide or protein may result from the operation of only one of these processes, or a combination of any of them.

The term "probe", as utilized herein, refers to nucleic acid sequences of variable
5 length, preferably between at least about 10 nucleotides (nt), about 100 nt, or as many as approximately, *e.g.*, 6,000 nt, depending upon the specific use. Probes are used in the detection of identical, similar, or complementary nucleic acid sequences. Longer length probes are generally obtained from a natural or recombinant source, are highly specific, and much slower to hybridize than shorter-length oligomer probes. Probes may be single-
10 stranded or double-stranded and designed to have specificity in PCR, membrane-based hybridization technologies, or ELISA-like technologies.

The term "isolated" nucleic acid molecule, as used herein, is a nucleic acid that is separated from other nucleic acid molecules which are present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank
15 the nucleic acid (*i.e.*, sequences located at the 5'- and 3'-termini of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. For example, in various embodiments, the isolated NOVX nucleic acid molecules can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotide sequences which naturally flank the nucleic acid molecule in genomic DNA of the cell/tissue from which the nucleic acid is
20 derived (*e.g.*, brain, heart, liver, spleen, *etc.*). Moreover, an "isolated" nucleic acid molecule, such as a cDNA molecule, can be substantially free of other cellular material, or culture medium, or of chemical precursors or other chemicals.

A nucleic acid molecule of the invention, *e.g.*, a nucleic acid molecule having the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or a
25 complement of this nucleotide sequence, can be isolated using standard molecular biology techniques and the sequence information provided herein. Using all or a portion of the nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, as a hybridization probe, NOVX molecules can be isolated using standard hybridization and cloning techniques (*e.g.*, as described in Sambrook, *et al.*, (eds.), MOLECULAR CLONING: A
30 LABORATORY MANUAL 2nd Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989; and Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993.)

A nucleic acid of the invention can be amplified using cDNA, mRNA or alternatively, genomic DNA, as a template with appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis. Furthermore,

5 oligonucleotides corresponding to NOVX nucleotide sequences can be prepared by standard synthetic techniques, *e.g.*, using an automated DNA synthesizer.

As used herein, the term "oligonucleotide" refers to a series of linked nucleotide residues. A short oligonucleotide sequence may be based on, or designed from, a genomic or cDNA sequence and is used to amplify, confirm, or reveal the presence of an identical,
10 similar or complementary DNA or RNA in a particular cell or tissue. Oligonucleotides comprise a nucleic acid sequence having about 10 nt, 50 nt, or 100 nt in length, preferably about 15 nt to 30 nt in length. In one embodiment of the invention, an oligonucleotide comprising a nucleic acid molecule less than 100 nt in length would further comprise at least 6 contiguous nucleotides of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or
15 a complement thereof. Oligonucleotides may be chemically synthesized and may also be used as probes.

In another embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule that is a complement of the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or a portion of this nucleotide sequence
20 (*e.g.*, a fragment that can be used as a probe or primer or a fragment encoding a biologically-active portion of a NOVX polypeptide). A nucleic acid molecule that is complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, is one that is sufficiently complementary to the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, that it can hydrogen bond with
25 few or no mismatches to the nucleotide sequence shown in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, thereby forming a stable duplex.

As used herein, the term "complementary" refers to Watson-Crick or Hoogsteen base pairing between nucleotides units of a nucleic acid molecule, and the term "binding" means the physical or chemical interaction between two polypeptides or compounds or associated
30 polypeptides or compounds or combinations thereof. Binding includes ionic, non-ionic, van der Waals, hydrophobic interactions, and the like. A physical interaction can be either direct or indirect. Indirect interactions may be through or due to the effects of another polypeptide or compound. Direct binding refers to interactions that do not take place through, or due to,

the effect of another polypeptide or compound, but instead are without other substantial chemical intermediates.

A "fragment" provided herein is defined as a sequence of at least 6 (contiguous) nucleic acids or at least 4 (contiguous) amino acids, a length sufficient to allow for specific hybridization in the case of nucleic acids or for specific recognition of an epitope in the case of amino acids, and is at most some portion less than a full length sequence. Fragments may be derived from any contiguous portion of a nucleic acid or amino acid sequence of choice.

A full-length NOVX clone is identified as containing an ATG translation start codon and an in-frame stop codon. Any disclosed NOVX nucleotide sequence lacking an ATG start codon therefore encodes a truncated C-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 5' direction of the disclosed sequence. Any disclosed NOVX nucleotide sequence lacking an in-frame stop codon similarly encodes a truncated N-terminal fragment of the respective NOVX polypeptide, and requires that the corresponding full-length cDNA extend in the 3' direction of the disclosed sequence.

A "derivative" is a nucleic acid sequence or amino acid sequence formed from the native compounds either directly, by modification or partial substitution. An "analog" is a nucleic acid sequence or amino acid sequence that has a structure similar to, but not identical to, the native compound, *e.g.* they differs from it in respect to certain components or side chains. Analogs may be synthetic or derived from a different evolutionary origin and may have a similar or opposite metabolic activity compared to wild type. A "homolog" is a nucleic acid sequence or amino acid sequence of a particular gene that is derived from different species.

Derivatives and analogs may be full length or other than full length. Derivatives or analogs of the nucleic acids or proteins of the invention include, but are not limited to, molecules comprising regions that are substantially homologous to the nucleic acids or proteins of the invention, in various embodiments, by at least about 70%, 80%, or 95% identity (with a preferred identity of 80-95%) over a nucleic acid or amino acid sequence of identical size or when compared to an aligned sequence in which the alignment is done by a computer homology program known in the art, or whose encoding nucleic acid is capable of hybridizing to the complement of a sequence encoding the proteins under stringent, moderately stringent, or low stringent conditions. *See e.g.* Ausubel, *et al.*, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, New York, NY, 1993, and below.

A "homologous nucleic acid sequence" or "homologous amino acid sequence," or variations thereof, refer to sequences characterized by a homology at the nucleotide level or amino acid level as discussed above. Homologous nucleotide sequences include those sequences coding for isoforms of NOVX polypeptides. Isoforms can be expressed in different tissues of the same organism as a result of, for example, alternative splicing of RNA. Alternatively, isoforms can be encoded by different genes. In the invention, homologous nucleotide sequences include nucleotide sequences encoding for a NOVX polypeptide of species other than humans, including, but not limited to: vertebrates, and thus can include, *e.g.*, frog, mouse, rat, rabbit, dog, cat, cow, horse, and other organisms.

Homologous nucleotide sequences also include, but are not limited to, naturally occurring allelic variations and mutations of the nucleotide sequences set forth herein. A homologous nucleotide sequence does not, however, include the exact nucleotide sequence encoding human NOVX protein. Homologous nucleic acid sequences include those nucleic acid sequences that encode conservative amino acid substitutions (see below) in SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, as well as a polypeptide possessing NOVX biological activity. Various biological activities of the NOVX proteins are described below.

A NOVX polypeptide is encoded by the open reading frame ("ORF") of a NOVX nucleic acid. An ORF corresponds to a nucleotide sequence that could potentially be translated into a polypeptide. A stretch of nucleic acids comprising an ORF is uninterrupted by a stop codon. An ORF that represents the coding sequence for a full protein begins with an ATG "start" codon and terminates with one of the three "stop" codons, namely, TAA, TAG, or TGA. For the purposes of this invention, an ORF may be any part of a coding sequence, with or without a start codon, a stop codon, or both. For an ORF to be considered as a good candidate for coding for a *bona fide* cellular protein, a minimum size requirement is often set, *e.g.*, a stretch of DNA that would encode a protein of 50 amino acids or more.

The nucleotide sequences determined from the cloning of the human NOVX genes allows for the generation of probes and primers designed for use in identifying and/or cloning NOVX homologues in other cell types, *e.g.* from other tissues, as well as NOVX homologues from other vertebrates. The probe/primer typically comprises substantially purified oligonucleotide. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 25, 50, 100, 150, 200, 250, 300, 350 or 400 consecutive sense strand nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172; or an anti-sense strand nucleotide sequence of SEQ ID

NO:2*n*-1, wherein *n* is an integer between 1 and 172; or of a naturally occurring mutant of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172.

Probes based on the human NOVX nucleotide sequences can be used to detect transcripts or genomic sequences encoding the same or homologous proteins. In various
5 embodiments, the probe has a detectable label attached, *e.g.* the label can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme co-factor. Such probes can be used as a part of a diagnostic test kit for identifying cells or tissues which mis-express a NOVX protein, such as by measuring a level of a NOVX-encoding nucleic acid in a sample of cells from a subject *e.g.*, detecting NOVX mRNA levels or determining whether a genomic NOVX
10 gene has been mutated or deleted.

"A polypeptide having a biologically-active portion of a NOVX polypeptide" refers to polypeptides exhibiting activity similar, but not necessarily identical to, an activity of a polypeptide of the invention, including mature forms, as measured in a particular biological assay, with or without dose dependency. A nucleic acid fragment encoding a
15 "biologically-active portion of NOVX" can be prepared by isolating a portion of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, that encodes a polypeptide having a NOVX biological activity (the biological activities of the NOVX proteins are described below), expressing the encoded portion of NOVX protein (*e.g.*, by recombinant expression *in vitro*) and assessing the activity of the encoded portion of NOVX.

20 NOVX Nucleic Acid and Polypeptide Variants

The invention further encompasses nucleic acid molecules that differ from the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, due to degeneracy of the genetic code and thus encode the same NOVX proteins as that encoded by the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172. In
25 another embodiment, an isolated nucleic acid molecule of the invention has a nucleotide sequence encoding a protein having an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172.

In addition to the human NOVX nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, it will be appreciated by those skilled in the art that DNA
30 sequence polymorphisms that lead to changes in the amino acid sequences of the NOVX polypeptides may exist within a population (*e.g.*, the human population). Such genetic polymorphism in the NOVX genes may exist among individuals within a population due to

natural allelic variation. As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules comprising an open reading frame (ORF) encoding a NOVX protein, preferably a vertebrate NOVX protein. Such natural allelic variations can typically result in 1-5% variance in the nucleotide sequence of the NOVX genes. Any and all such nucleotide variations and resulting amino acid polymorphisms in the NOVX polypeptides, which are the result of natural allelic variation and that do not alter the functional activity of the NOVX polypeptides, are intended to be within the scope of the invention.

Moreover, nucleic acid molecules encoding NOVX proteins from other species, and thus that have a nucleotide sequence that differs from a human SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, are intended to be within the scope of the invention. Nucleic acid molecules corresponding to natural allelic variants and homologues of the NOVX cDNAs of the invention can be isolated based on their homology to the human NOVX nucleic acids disclosed herein using the human cDNAs, or a portion thereof, as a hybridization probe according to standard hybridization techniques under stringent hybridization conditions.

Accordingly, in another embodiment, an isolated nucleic acid molecule of the invention is at least 6 nucleotides in length and hybridizes under stringent conditions to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172. In another embodiment, the nucleic acid is at least 10, 25, 50, 100, 250, 500, 750, 1000, 1500, or 2000 or more nucleotides in length. In yet another embodiment, an isolated nucleic acid molecule of the invention hybridizes to the coding region. As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 65% homologous to each other typically remain hybridized to each other.

Homologs (*i.e.*, nucleic acids encoding NOVX proteins derived from species other than human) or other related sequences (*e.g.*, paralogs) can be obtained by low, moderate or high stringency hybridization with all or a portion of the particular human sequence as a probe using methods well known in the art for nucleic acid hybridization and cloning.

As used herein, the phrase "stringent hybridization conditions" refers to conditions under which a probe, primer or oligonucleotide will hybridize to its target sequence, but to no other sequences. Stringent conditions are sequence-dependent and will be different in different circumstances. Longer sequences hybridize specifically at higher temperatures than shorter sequences. Generally, stringent conditions are selected to be about 5 °C lower than the

thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength, pH and nucleic acid concentration) at which 50% of the probes complementary to the target sequence hybridize to the target sequence at equilibrium. Since the target sequences are generally present at excess, at T_m , 50% of the probes are occupied at equilibrium. Typically, stringent conditions will be those in which the salt concentration is less than about 1.0 M sodium ion, typically about 0.01 to 1.0 M sodium ion (or other salts) at pH 7.0 to 8.3 and the temperature is at least about 30 °C for short probes, primers or oligonucleotides (e.g., 10 nt to 50 nt) and at least about 60 °C for longer probes, primers and oligonucleotides. Stringent conditions may also be achieved with the addition of destabilizing agents, such as formamide.

Stringent conditions are known to those skilled in the art and can be found in Ausubel, *et al.*, (eds.), CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. Preferably, the conditions are such that sequences at least about 65%, 70%, 75%, 85%, 90%, 95%, 98%, or 99% homologous to each other typically remain hybridized to each other. A non-limiting example of stringent hybridization conditions are hybridization in a high salt buffer comprising 6X SSC, 50 mM Tris-HCl (pH 7.5), 1 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.02% BSA, and 500 mg/ml denatured salmon sperm DNA at 65°C, followed by one or more washes in 0.2X SSC, 0.01% BSA at 50°C. An isolated nucleic acid molecule of the invention that hybridizes under stringent conditions to a sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, corresponds to a naturally-occurring nucleic acid molecule. As used herein, a "naturally-occurring" nucleic acid molecule refers to an RNA or DNA molecule having a nucleotide sequence that occurs in nature (e.g., encodes a natural protein).

In a second embodiment, a nucleic acid sequence that is hybridizable to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or fragments, analogs or derivatives thereof, under conditions of moderate stringency is provided. A non-limiting example of moderate stringency hybridization conditions are hybridization in 6X SSC, 5X Reinhardt's solution, 0.5% SDS and 100 mg/ml denatured salmon sperm DNA at 55 °C, followed by one or more washes in 1X SSC, 0.1% SDS at 37 °C. Other conditions of moderate stringency that may be used are well-known within the art. *See, e.g.*, Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Krieger, 1990; GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY.

In a third embodiment, a nucleic acid that is hybridizable to the nucleic acid molecule comprising the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or fragments, analogs or derivatives thereof, under conditions of low stringency, is provided. A non-limiting example of low stringency hybridization conditions are

5 hybridization in 35% formamide, 5X SSC, 50 mM Tris-HCl (pH 7.5), 5 mM EDTA, 0.02% PVP, 0.02% Ficoll, 0.2% BSA, 100 mg/ml denatured salmon sperm DNA, 10% (wt/vol) dextran sulfate at 40°C, followed by one or more washes in 2X SSC, 25 mM Tris-HCl (pH 7.4), 5 mM EDTA, and 0.1% SDS at 50°C. Other conditions of low stringency that may be used are well known in the art (*e.g.*, as employed for cross-species hybridizations). *See, e.g.*,

10 Ausubel, *et al.* (eds.), 1993, CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, NY, and Kriegler, 1990, GENE TRANSFER AND EXPRESSION, A LABORATORY MANUAL, Stockton Press, NY; Shilo and Weinberg, 1981. *Proc Natl Acad Sci USA* 78: 6789-6792.

Conservative Mutations

In addition to naturally-occurring allelic variants of NOVX sequences that may exist

15 in the population, the skilled artisan will further appreciate that changes can be introduced by mutation into the nucleotide sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, thereby leading to changes in the amino acid sequences of the encoded NOVX protein, without altering the functional ability of that NOVX protein. For example, nucleotide substitutions leading to amino acid substitutions at "non-essential" amino acid

20 residues can be made in the sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172. A "non-essential" amino acid residue is a residue that can be altered from the wild-type sequences of the NOVX proteins without altering their biological activity, whereas an "essential" amino acid residue is required for such biological activity. For example, amino acid residues that are conserved among the NOVX proteins of the invention are predicted to

25 be particularly non-amenable to alteration. Amino acids for which conservative substitutions can be made are well-known within the art.

Another aspect of the invention pertains to nucleic acid molecules encoding NOVX proteins that contain changes in amino acid residues that are not essential for activity. Such NOVX proteins differ in amino acid sequence from SEQ ID NO:2*n*-1, wherein *n* is an integer

30 between 1 and 172, yet retain biological activity. In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises an amino acid sequence at least about 40% homologous to the amino acid

sequences of SEQ ID NO:2 n , wherein n is an integer between 1 and 172. Preferably, the protein encoded by the nucleic acid molecule is at least about 60% homologous to SEQ ID NO:2 n , wherein n is an integer between 1 and 172; more preferably at least about 70% homologous to SEQ ID NO:2 n , wherein n is an integer between 1 and 172; still more preferably at least about 80% homologous to SEQ ID NO:2 n , wherein n is an integer between 1 and 172; even more preferably at least about 90% homologous to SEQ ID NO:2 n , wherein n is an integer between 1 and 172; and most preferably at least about 95% homologous to SEQ ID NO:2 n , wherein n is an integer between 1 and 172.

An isolated nucleic acid molecule encoding a NOVX protein homologous to the protein of SEQ ID NO:2 n , wherein n is an integer between 1 and 172, can be created by introducing one or more nucleotide substitutions, additions or deletions into the nucleotide sequence of SEQ ID NO:2 n -1, wherein n is an integer between 1 and 172, such that one or more amino acid substitutions, additions or deletions are introduced into the encoded protein.

Mutations can be introduced any one of SEQ ID NO:2 n -1, wherein n is an integer between 1 and 172, by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis. Preferably, conservative amino acid substitutions are made at one or more predicted, non-essential amino acid residues. A "conservative amino acid substitution" is one in which the amino acid residue is replaced with an amino acid residue having a similar side chain. Families of amino acid residues having similar side chains have been defined within the art. These families include amino acids with basic side chains (*e.g.*, lysine, arginine, histidine), acidic side chains (*e.g.*, aspartic acid, glutamic acid), uncharged polar side chains (*e.g.*, glycine, asparagine, glutamine, serine, threonine, tyrosine, cysteine), nonpolar side chains (*e.g.*, alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (*e.g.*, threonine, valine, isoleucine) and aromatic side chains (*e.g.*, tyrosine, phenylalanine, tryptophan, histidine). Thus, a predicted non-essential amino acid residue in the NOVX protein is replaced with another amino acid residue from the same side chain family. Alternatively, in another embodiment, mutations can be introduced randomly along all or part of a NOVX coding sequence, such as by saturation mutagenesis, and the resultant mutants can be screened for NOVX biological activity to identify mutants that retain activity. Following mutagenesis of a nucleic acid of SEQ ID NO:2 n -1, wherein n is an integer between 1 and 172, the encoded protein can be expressed by any recombinant technology known in the art and the activity of the protein can be determined.

The relatedness of amino acid families may also be determined based on side chain interactions. Substituted amino acids may be fully conserved “strong” residues or fully conserved “weak” residues. The “strong” group of conserved amino acid residues may be any one of the following groups: STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY.

5 FYW, wherein the single letter amino acid codes are grouped by those amino acids that may be substituted for each other. Likewise, the “weak” group of conserved residues may be any one of the following: CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, HFY, wherein the letters within each group represent the single letter amino acid code.

10 In one embodiment, a mutant NOVX protein can be assayed for (i) the ability to form protein:protein interactions with other NOVX proteins, other cell-surface proteins, or biologically-active portions thereof, (ii) complex formation between a mutant NOVX protein and a NOVX ligand; or (iii) the ability of a mutant NOVX protein to bind to an intracellular target protein or biologically-active portion thereof; (e.g. avidin proteins).

15 In yet another embodiment, a mutant NOVX protein can be assayed for the ability to regulate a specific biological function (e.g., regulation of insulin release).

Interfering RNA

In one aspect of the invention, NOVX gene expression can be attenuated by RNA interference. One approach well-known in the art is short interfering RNA (siRNA) mediated gene silencing where expression products of a NOVX gene are targeted by specific double stranded NOVX derived siRNA nucleotide sequences that are complementary to at least a 19-25 nt long segment of the NOVX gene transcript, including the 5' untranslated (UT) region, the ORF, or the 3' UT region. See, e.g., PCT applications WO00/44895, WO99/32619, WO01/75164, WO01/92513, WO 01/29058, WO01/89304, WO02/16620, and WO02/29858, each incorporated by reference herein in their entirety. Targeted genes can be a NOVX gene, or an upstream or downstream modulator of the NOVX gene. Nonlimiting examples of upstream or downstream modulators of a NOVX gene include, e.g., a transcription factor that binds the NOVX gene promoter, a kinase or phosphatase that interacts with a NOVX polypeptide, and polypeptides involved in a NOVX regulatory pathway.

30 According to the methods of the present invention, NOVX gene expression is silenced using short interfering RNA. A NOVX polynucleotide according to the invention includes a siRNA polynucleotide. Such a NOVX siRNA can be obtained using a NOVX polynucleotide

sequence, for example, by processing the NOVX ribopolynucleotide sequence in a cell-free system, such as but not limited to a *Drosophila* extract, or by transcription of recombinant double stranded NOVX RNA or by chemical synthesis of nucleotide sequences homologous to a NOVX sequence. *See, e.g.*, Tuschl, Zamore, Lehmann, Bartel and Sharp (1999), *Genes & Dev.* 13: 3191-3197, incorporated herein by reference in its entirety. When synthesized, a typical 0.2 micromolar-scale RNA synthesis provides about 1 milligram of siRNA, which is sufficient for 1000 transfection experiments using a 24-well tissue culture plate format.

The most efficient silencing is generally observed with siRNA duplexes composed of a 21-nt sense strand and a 21-nt antisense strand, paired in a manner to have a 2-nt 3' overhang. The sequence of the 2-nt 3' overhang makes an additional small contribution to the specificity of siRNA target recognition. The contribution to specificity is localized to the unpaired nucleotide adjacent to the first paired bases. In one embodiment, the nucleotides in the 3' overhang are ribonucleotides. In an alternative embodiment, the nucleotides in the 3' overhang are deoxyribonucleotides. Using 2'-deoxyribonucleotides in the 3' overhangs is as efficient as using ribonucleotides, but deoxyribonucleotides are often cheaper to synthesize and are most likely more nuclease resistant.

A contemplated recombinant expression vector of the invention comprises a NOVX DNA molecule cloned into an expression vector comprising operatively-linked regulatory sequences flanking the NOVX sequence in a manner that allows for expression (by transcription of the DNA molecule) of both strands. An RNA molecule that is antisense to NOVX mRNA is transcribed by a first promoter (*e.g.*, a promoter sequence 3' of the cloned DNA) and an RNA molecule that is the sense strand for the NOVX mRNA is transcribed by a second promoter (*e.g.*, a promoter sequence 5' of the cloned DNA). The sense and antisense strands may hybridize *in vivo* to generate siRNA constructs for silencing of the NOVX gene. Alternatively, two constructs can be utilized to create the sense and anti-sense strands of a siRNA construct. Finally, cloned DNA can encode a construct having secondary structure, wherein a single transcript has both the sense and complementary antisense sequences from the target gene or genes. In an example of this embodiment, a hairpin RNAi product is homologous to all or a portion of the target gene. In another example, a hairpin RNAi product is a siRNA. The regulatory sequences flanking the NOVX sequence may be identical or may be different, such that their expression may be modulated independently, or in a temporal or spatial manner.

In a specific embodiment, siRNAs are transcribed intracellularly by cloning the NOVX gene templates into a vector containing, *e.g.*, a RNA pol III transcription unit from the smaller nuclear RNA (snRNA) U6 or the human RNase P RNA H1. One example of a vector system is the GeneSuppressorTM RNA Interference kit (commercially available from Imgenex). The U6 and H1 promoters are members of the type III class of Pol III promoters. The +1 nucleotide of the U6-like promoters is always guanosine, whereas the +1 for H1 promoters is adenosine. The termination signal for these promoters is defined by five consecutive thymidines. The transcript is typically cleaved after the second uridine. Cleavage at this position generates a 3' UU overhang in the expressed siRNA, which is similar to the 3' overhangs of synthetic siRNAs. Any sequence less than 400 nucleotides in length can be transcribed by these promoter, therefore they are ideally suited for the expression of around 21-nucleotide siRNAs in, *e.g.*, an approximately 50-nucleotide RNA stem-loop transcript.

A siRNA vector appears to have an advantage over synthetic siRNAs where long term knock-down of expression is desired. Cells transfected with a siRNA expression vector would experience steady, long-term mRNA inhibition. In contrast, cells transfected with exogenous synthetic siRNAs typically recover from mRNA suppression within seven days or ten rounds of cell division. The long-term gene silencing ability of siRNA expression vectors may provide for applications in gene therapy.

In general, siRNAs are chopped from longer dsRNA by an ATP-dependent ribonuclease called DICER. DICER is a member of the RNase III family of double-stranded RNA-specific endonucleases. The siRNAs assemble with cellular proteins into an endonuclease complex. *In vitro* studies in *Drosophila* suggest that the siRNAs/protein complex (siRNP) is then transferred to a second enzyme complex, called an RNA-induced silencing complex (RISC), which contains an endoribonuclease that is distinct from DICER. RISC uses the sequence encoded by the antisense siRNA strand to find and destroy mRNAs of complementary sequence. The siRNA thus acts as a guide, restricting the ribonuclease to cleave only mRNAs complementary to one of the two siRNA strands.

A NOVX mRNA region to be targeted by siRNA is generally selected from a desired NOVX sequence beginning 50 to 100 nt downstream of the start codon. Alternatively, 5' or 3' UTRs and regions nearby the start codon can be used but are generally avoided, as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP or RISC endonuclease complex. An initial BLAST homology search for the selected siRNA sequence is done

against an available nucleotide sequence library to ensure that only one gene is targeted. Specificity of target recognition by siRNA duplexes indicate that a single point mutation located in the paired region of an siRNA duplex is sufficient to abolish target mRNA degradation. See, Elbashir *et al.* 2001 EMBO J. 20(23):6877-88. Hence, consideration
5 should be taken to accommodate SNPs, polymorphisms, allelic variants or species-specific variations when targeting a desired gene.

In one embodiment, a complete NOVX siRNA experiment includes the proper negative control. A negative control siRNA generally has the same nucleotide composition as the NOVX siRNA but lack significant sequence homology to the genome. Typically, one
10 would scramble the nucleotide sequence of the NOVX siRNA and do a homology search to make sure it lacks homology to any other gene.

Two independent NOVX siRNA duplexes can be used to knock-down a target NOVX gene. This helps to control for specificity of the silencing effect. In addition, expression of two independent genes can be simultaneously knocked down by using equal concentrations
15 of different NOVX siRNA duplexes, *e.g.*, a NOVX siRNA and an siRNA for a regulator of a NOVX gene or polypeptide. Availability of siRNA-associating proteins is believed to be more limiting than target mRNA accessibility.

A targeted NOVX region is typically a sequence of two adenines (AA) and two thymidines (TT) divided by a spacer region of nineteen (N19) residues (*e.g.*, AA(N19)TT).

20 A desirable spacer region has a G/C-content of approximately 30% to 70%, and more preferably of about 50%. If the sequence AA(N19)TT is not present in the target sequence, an alternative target region would be AA(N21). The sequence of the NOVX sense siRNA corresponds to (N19)TT or N21, respectively. In the latter case, conversion of the 3' end of the sense siRNA to TT can be performed if such a sequence does not naturally occur in the
25 NOVX polynucleotide. The rationale for this sequence conversion is to generate a symmetric duplex with respect to the sequence composition of the sense and antisense 3' overhangs. Symmetric 3' overhangs may help to ensure that the siRNPs are formed with approximately equal ratios of sense and antisense target RNA-cleaving siRNPs. See, *e.g.*, Elbashir, Lendeckel and Tuschl (2001). Genes & Dev. 15: 188-200, incorporated by reference herein in
30 its entirety. The modification of the overhang of the sense sequence of the siRNA duplex is not expected to affect targeted mRNA recognition, as the antisense siRNA strand guides target recognition.

Alternatively, if the NOVX target mRNA does not contain a suitable AA(N21) sequence, one may search for the sequence NA(N21). Further, the sequence of the sense strand and antisense strand may still be synthesized as 5' (N19)TT, as it is believed that the sequence of the 3'-most nucleotide of the antisense siRNA does not contribute to specificity.

5 Unlike antisense or ribozyme technology, the secondary structure of the target mRNA does not appear to have a strong effect on silencing. *See, Harborth, et al. (2001) J. Cell Science 114: 4557-4565, incorporated by reference in its entirety.*

Transfection of NOVX siRNA duplexes can be achieved using standard nucleic acid transfection methods, for example, OLIGOFECTAMINE Reagent (commercially available
10 from Invitrogen). An assay for NOVX gene silencing is generally performed approximately 2 days after transfection. No NOVX gene silencing has been observed in the absence of transfection reagent, allowing for a comparative analysis of the wild-type and silenced NOVX phenotypes. In a specific embodiment, for one well of a 24-well plate, approximately 0.84 μ g of the siRNA duplex is generally sufficient. Cells are typically seeded the previous
15 day, and are transfected at about 50% confluence. The choice of cell culture media and conditions are routine to those of skill in the art, and will vary with the choice of cell type. The efficiency of transfection may depend on the cell type, but also on the passage number and the confluency of the cells. The time and the manner of formation of siRNA-liposome complexes (*e.g.* inversion versus vortexing) are also critical. Low transfection efficiencies
20 are the most frequent cause of unsuccessful NOVX silencing. The efficiency of transfection needs to be carefully examined for each new cell line to be used. Preferred cell are derived from a mammal, more preferably from a rodent such as a rat or mouse, and most preferably from a human. Where used for therapeutic treatment, the cells are preferentially autologous, although non-autologous cell sources are also contemplated as within the scope of the present
25 invention.

For a control experiment, transfection of 0.84 μ g single-stranded sense NOVX siRNA will have no effect on NOVX silencing, and 0.84 μ g antisense siRNA has a weak silencing effect when compared to 0.84 μ g of duplex siRNAs. Control experiments again allow for a comparative analysis of the wild-type and silenced NOVX phenotypes. To control for
30 transfection efficiency, targeting of common proteins is typically performed, for example targeting of lamin A/C or transfection of a CMV-driven EGFP-expression plasmid (*e.g.* commercially available from Clontech). In the above example, a determination of the fraction of lamin A/C knockdown in cells is determined the next day by such techniques as

immunofluorescence, Western blot, Northern blot or other similar assays for protein expression or gene expression. Lamin A/C monoclonal antibodies may be obtained from Santa Cruz Biotechnology.

Depending on the abundance and the half life (or turnover) of the targeted NOVX polynucleotide in a cell, a knock-down phenotype may become apparent after 1 to 3 days, or even later. In cases where no NOVX knock-down phenotype is observed, depletion of the NOVX polynucleotide may be observed by immunofluorescence or Western blotting. If the NOVX polynucleotide is still abundant after 3 days, cells need to be split and transferred to a fresh 24-well plate for re-transfection. If no knock-down of the targeted protein is observed, it may be desirable to analyze whether the target mRNA (NOVX or a NOVX upstream or downstream gene) was effectively destroyed by the transfected siRNA duplex. Two days after transfection, total RNA is prepared, reverse transcribed using a target-specific primer, and PCR-amplified with a primer pair covering at least one exon-exon junction in order to control for amplification of pre-mRNAs. RT/PCR of a non-targeted mRNA is also needed as control. Effective depletion of the mRNA yet undetectable reduction of target protein may indicate that a large reservoir of stable NOVX protein may exist in the cell. Multiple transfection in sufficiently long intervals may be necessary until the target protein is finally depleted to a point where a phenotype may become apparent. If multiple transfection steps are required, cells are split 2 to 3 days after transfection. The cells may be transfected immediately after splitting.

An inventive therapeutic method of the invention contemplates administering a NOVX siRNA construct as therapy to compensate for increased or aberrant NOVX expression or activity. The NOVX ribopolynucleotide is obtained and processed into siRNA fragments, or a NOVX siRNA is synthesized, as described above. The NOVX siRNA is administered to cells or tissues using known nucleic acid transfection techniques, as described above. A NOVX siRNA specific for a NOVX gene will decrease or knockdown NOVX transcription products, which will lead to reduced NOVX polypeptide production, resulting in reduced NOVX polypeptide activity in the cells or tissues.

The present invention also encompasses a method of treating a disease or condition associated with the presence of a NOVX protein in an individual comprising administering to the individual an RNAi construct that targets the mRNA of the protein (the mRNA that encodes the protein) for degradation. A specific RNAi construct includes a siRNA or a double stranded gene transcript that is processed into siRNAs. Upon treatment, the target

protein is not produced or is not produced to the extent it would be in the absence of the treatment.

Where the NOVX gene function is not correlated with a known phenotype, a control sample of cells or tissues from healthy individuals provides a reference standard for determining NOVX expression levels. Expression levels are detected using the assays described, *e.g.*, RT-PCR, Northern blotting, Western blotting, ELISA, and the like. A subject sample of cells or tissues is taken from a mammal, preferably a human subject, suffering from a disease state. The NOVX ribopolynucleotide is used to produce siRNA constructs, that are specific for the NOVX gene product. These cells or tissues are treated by administering NOVX siRNA's to the cells or tissues by methods described for the transfection of nucleic acids into a cell or tissue, and a change in NOVX polypeptide or polynucleotide expression is observed in the subject sample relative to the control sample, using the assays described. This NOVX gene knockdown approach provides a rapid method for determination of a NOVX minus (NOVX⁻) phenotype in the treated subject sample. The NOVX⁻ phenotype observed in the treated subject sample thus serves as a marker for monitoring the course of a disease state during treatment.

In specific embodiments, a NOVX siRNA is used in therapy. Methods for the generation and use of a NOVX siRNA are known to those skilled in the art. Example techniques are provided below.

Production of RNAs

Sense RNA (ssRNA) and antisense RNA (asRNA) of NOVX are produced using known methods such as transcription in RNA expression vectors. In the initial experiments, the sense and antisense RNA are about 500 bases in length each. The produced ssRNA and asRNA (0.5 μ M) in 10 mM Tris-HCl (pH 7.5) with 20 mM NaCl were heated to 95° C for 1 min then cooled and annealed at room temperature for 12 to 16 h. The RNAs are precipitated and resuspended in lysis buffer (below). To monitor annealing, RNAs are electrophoresed in a 2% agarose gel in TBE buffer and stained with ethidium bromide. See, *e.g.*, Sambrook et al., Molecular Cloning. Cold Spring Harbor Laboratory Press, Plainview, N.Y. (1989).

Lysate Preparation

Untreated rabbit reticulocyte lysate (Ambion) are assembled according to the manufacturer's directions. dsRNA is incubated in the lysate at 30° C for 10 min prior to the addition of mRNAs. Then NOVX mRNAs are added and the incubation continued for an

additional 60 min. The molar ratio of double stranded RNA and mRNA is about 200:1. The NOVX mRNA is radiolabeled (using known techniques) and its stability is monitored by gel electrophoresis.

In a parallel experiment made with the same conditions, the double stranded RNA is internally radiolabeled with a ^{32}P -ATP. Reactions are stopped by the addition of 2 X proteinase K buffer and deproteinized as described previously (Tuschl *et al.*, Genes Dev., 13:3191-3197 (1999)). Products are analyzed by electrophoresis in 15% or 18% polyacrylamide sequencing gels using appropriate RNA standards. By monitoring the gels for radioactivity, the natural production of 10 to 25 nt RNAs from the double stranded RNA can be determined.

The band of double stranded RNA, about 21-23 bps, is eluted. The efficacy of these 21-23 mers for suppressing NOVX transcription is assayed in vitro using the same rabbit reticulocyte assay described above using 50 nanomolar of double stranded 21-23 mer for each assay. The sequence of these 21-23 mers is then determined using standard nucleic acid sequencing techniques.

RNA Preparation

21 nt RNAs, based on the sequence determined above, are chemically synthesized using Expedite RNA phosphoramidites and thymidine phosphoramidite (Proligo, Germany). Synthetic oligonucleotides are deprotected and gel-purified (Elbashir, Lendeckel, & Tuschl, Genes & Dev. 15, 188-200 (2001)), followed by Sep-Pak C18 cartridge (Waters, Milford, Mass., USA) purification (Tuschl, et al., Biochemistry, 32:11658-11668 (1993)).

These RNAs (20 μM) single strands are incubated in annealing buffer (100 mM potassium acetate, 30 mM HEPES-KOH at pH 7.4, 2 mM magnesium acetate) for 1 min at 90° C followed by 1 h at 37° C.

Cell Culture

A cell culture known in the art to regularly express NOVX is propagated using standard conditions. 24 hours before transfection, at approx. 80% confluency, the cells are trypsinized and diluted 1:5 with fresh medium without antibiotics ($1-3 \times 10^5$ cells/ml) and transferred to 24-well plates (500 ml/well). Transfection is performed using a commercially available lipofection kit and NOVX expression is monitored using standard techniques with positive and negative control. A positive control is cells that naturally express NOVX while a negative control is cells that do not express NOVX. Base-paired 21 and 22 nt siRNAs with

overhanging 3' ends mediate efficient sequence-specific mRNA degradation in lysates and in cell culture. Different concentrations of siRNAs are used. An efficient concentration for suppression in vitro in mammalian culture is between 25 nM to 100 nM final concentration. This indicates that siRNAs are effective at concentrations that are several orders of magnitude below the concentrations applied in conventional antisense or ribozyme gene targeting experiments.

The above method provides a way both for the deduction of NOVX siRNA sequence and the use of such siRNA for in vitro suppression. In vivo suppression may be performed using the same siRNA using well known in vivo transfection or gene therapy transfection techniques.

Antisense Nucleic Acids

Another aspect of the invention pertains to isolated antisense nucleic acid molecules that are hybridizable to or complementary to the nucleic acid molecule comprising the nucleotide sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or fragments, analogs or derivatives thereof. An "antisense" nucleic acid comprises a nucleotide sequence that is complementary to a "sense" nucleic acid encoding a protein (*e.g.*, complementary to the coding strand of a double-stranded cDNA molecule or complementary to an mRNA sequence). In specific aspects, antisense nucleic acid molecules are provided that comprise a sequence complementary to at least about 10, 25, 50, 100, 250 or 500 nucleotides or an entire NOVX coding strand, or to only a portion thereof. Nucleic acid molecules encoding fragments, homologs, derivatives and analogs of a NOVX protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172, or antisense nucleic acids complementary to a NOVX nucleic acid sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, are additionally provided.

In one embodiment, an antisense nucleic acid molecule is antisense to a "coding region" of the coding strand of a nucleotide sequence encoding a NOVX protein. The term "coding region" refers to the region of the nucleotide sequence comprising codons which are translated into amino acid residues. In another embodiment, the antisense nucleic acid molecule is antisense to a "noncoding region" of the coding strand of a nucleotide sequence encoding the NOVX protein. The term "noncoding region" refers to 5' and 3' sequences which flank the coding region that are not translated into amino acids (*i.e.*, also referred to as 5' and 3' untranslated regions).

Given the coding strand sequences encoding the NOVX protein disclosed herein, antisense nucleic acids of the invention can be designed according to the rules of Watson and Crick or Hoogsteen base pairing. The antisense nucleic acid molecule can be complementary to the entire coding region of NOVX mRNA, but more preferably is an oligonucleotide that is antisense to only a portion of the coding or noncoding region of NOVX mRNA. For example, the antisense oligonucleotide can be complementary to the region surrounding the translation start site of NOVX mRNA. An antisense oligonucleotide can be, for example, about 5, 10, 15, 20, 25, 30, 35, 40, 45 or 50 nucleotides in length. An antisense nucleic acid of the invention can be constructed using chemical synthesis or enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally-occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids (*e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used).

Examples of modified nucleotides that can be used to generate the antisense nucleic acid include: 5-fluorouracil, 5-bromouracil, 5-chlorouracil, 5-iodouracil, hypoxanthine, xanthine, 4-acetylcytosine, 5-carboxymethylaminomethyl-2-thiouridine, 5-(carboxyhydroxymethyl) uracil, 5-carboxymethylaminomethyluracil, dihydrouracil, beta-D-galactosylqueosine, inosine, N6-isopentenyladenine, 1-methylguanine, 1-methylinosine, 2,2-dimethylguanine, 2-methyladenine, 2-methylguanine, 5-methoxyuracil, 3-methylcytosine, 5-methylcytosine, N6-adenine, 7-methylguanine, 5-methylaminomethyluracil, 5-methoxyaminomethyl-2-thiouracil, 2-thiouracil, 4-thiouracil, beta-D-mannosylqueosine, 5'-methoxycarboxymethyluracil, 2-methylthio-N6-isopentenyladenine, uracil-5-oxyacetic acid (*v*), wybutoxosine, pseudouracil, queosine, 2-thiocytosine, 5-methyl-2-thiouracil, 5-methyluracil, uracil-5-oxyacetic acid methylester, uracil-5-oxyacetic acid (*v*), 5-methyl-2-thiouracil, 3-(3-amino-3-N-2-carboxypropyl) uracil, (acp3)*w*, and 2,6-diaminopurine. Alternatively, the antisense nucleic acid can be produced biologically using an expression vector into which a nucleic acid has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid will be of an antisense orientation to a target nucleic acid of interest, described further in the following subsection).

The antisense nucleic acid molecules of the invention are typically administered to a subject or generated *in situ* such that they hybridize with or bind to cellular mRNA and/or genomic DNA encoding a NOVX protein to thereby inhibit expression of the protein (*e.g.*, by inhibiting transcription and/or translation). The hybridization can be by conventional

5 nucleotide complementarity to form a stable duplex, or, for example, in the case of an antisense nucleic acid molecule that binds to DNA duplexes, through specific interactions in the major groove of the double helix. An example of a route of administration of antisense nucleic acid molecules of the invention includes direct injection at a tissue site.

Alternatively, antisense nucleic acid molecules can be modified to target selected cells and

10 then administered systemically. For example, for systemic administration, antisense molecules can be modified such that they specifically bind to receptors or antigens expressed on a selected cell surface (*e.g.*, by linking the antisense nucleic acid molecules to peptides or antibodies that bind to cell surface receptors or antigens). The antisense nucleic acid molecules can also be delivered to cells using the vectors described herein. To achieve

15 sufficient nucleic acid molecules, vector constructs in which the antisense nucleic acid molecule is placed under the control of a strong pol II or pol III promoter are preferred.

In yet another embodiment, the antisense nucleic acid molecule of the invention is an α -anomeric nucleic acid molecule. An α -anomeric nucleic acid molecule forms specific double-stranded hybrids with complementary RNA in which, contrary to the usual β -units,

20 the strands run parallel to each other. *See, e.g., Gaultier, et al., 1987. Nucl. Acids Res. 15: 6625-6641.* The antisense nucleic acid molecule can also comprise a 2'-o-methylribonucleotide (*See, e.g., Inoue, et al. 1987. Nucl. Acids Res. 15: 6131-6148*) or a chimeric RNA-DNA analogue (*See, e.g., Inoue, et al., 1987. FEBS Lett. 215: 327-330.*

Ribozymes and PNA Moieties

25 Nucleic acid modifications include, by way of non-limiting example, modified bases, and nucleic acids whose sugar phosphate backbones are modified or derivatized. These modifications are carried out at least in part to enhance the chemical stability of the modified nucleic acid, such that they may be used, for example, as antisense binding nucleic acids in therapeutic applications in a subject.

30 In one embodiment, an antisense nucleic acid of the invention is a ribozyme. Ribozymes are catalytic RNA molecules with ribonuclease activity that are capable of cleaving a single-stranded nucleic acid, such as an mRNA, to which they have a

complementary region. Thus, ribozymes (*e.g.*, hammerhead ribozymes as described in Haselhoff and Gerlach 1988. *Nature* 334: 585-591) can be used to catalytically cleave NOVX mRNA transcripts to thereby inhibit translation of NOVX mRNA. A ribozyme having specificity for a NOVX-encoding nucleic acid can be designed based upon the nucleotide
5 sequence of a NOVX cDNA disclosed herein (*i.e.*, SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172). For example, a derivative of a *Tetrahymena* L-19 IVS RNA can be constructed in which the nucleotide sequence of the active site is complementary to the nucleotide sequence to be cleaved in a NOVX-encoding mRNA. *See, e.g.*, U.S. Patent 4,987,071 to Cech, *et al.* and U.S. Patent 5,116,742 to Cech, *et al.* NOVX mRNA can also be
10 used to select a catalytic RNA having a specific ribonuclease activity from a pool of RNA molecules. *See, e.g.*, Bartel *et al.*, (1993) *Science* 261:1411-1418.

Alternatively, NOVX gene expression can be inhibited by targeting nucleotide sequences complementary to the regulatory region of the NOVX nucleic acid (*e.g.*, the NOVX promoter and/or enhancers) to form triple helical structures that prevent transcription
15 of the NOVX gene in target cells. *See, e.g.*, Helene, 1991. *Anticancer Drug Des.* 6: 569-84; Helene, *et al.* 1992. *Ann. N.Y. Acad. Sci.* 660: 27-36; Maher, 1992. *Bioassays* 14: 807-15.

In various embodiments, the NOVX nucleic acids can be modified at the base moiety, sugar moiety or phosphate backbone to improve, *e.g.*, the stability, hybridization, or solubility of the molecule. For example, the deoxyribose phosphate backbone of the nucleic acids can
20 be modified to generate peptide nucleic acids. *See, e.g.*, Hyrup, *et al.*, 1996. *Bioorg Med Chem* 4: 5-23. As used herein, the terms "peptide nucleic acids" or "PNAs" refer to nucleic acid mimics (*e.g.*, DNA mimics) in which the deoxyribose phosphate backbone is replaced by a pseudopeptide backbone and only the four natural nucleotide bases are retained. The neutral backbone of PNAs has been shown to allow for specific hybridization to DNA and
25 RNA under conditions of low ionic strength. The synthesis of PNA oligomer can be performed using standard solid phase peptide synthesis protocols as described in Hyrup, *et al.*, 1996. *supra*; Perry-O'Keefe, *et al.*, 1996. *Proc. Natl. Acad. Sci. USA* 93: 14670-14675.

PNAs of NOVX can be used in therapeutic and diagnostic applications. For example, PNAs can be used as antisense or antigene agents for sequence-specific modulation of gene
30 expression by, *e.g.*, inducing transcription or translation arrest or inhibiting replication. PNAs of NOVX can also be used, for example, in the analysis of single base pair mutations in a gene (*e.g.*, PNA directed PCR clamping; as artificial restriction enzymes when used in combination with other enzymes, *e.g.*, S₁ nucleases (*See, Hyrup, et al.*, 1996.*supra*); or as

probes or primers for DNA sequence and hybridization (*See, Hyrup, et al., 1996, supra; Perry-O'Keefe, et al., 1996. supra*).

In another embodiment, PNAs of NOVX can be modified, *e.g.*, to enhance their stability or cellular uptake, by attaching lipophilic or other helper groups to PNA, by the formation of PNA-DNA chimeras, or by the use of liposomes or other techniques of drug delivery known in the art. For example, PNA-DNA chimeras of NOVX can be generated that may combine the advantageous properties of PNA and DNA. Such chimeras allow DNA recognition enzymes (*e.g.*, RNase H and DNA polymerases) to interact with the DNA portion while the PNA portion would provide high binding affinity and specificity. PNA-DNA chimeras can be linked using linkers of appropriate lengths selected in terms of base stacking, number of bonds between the nucleotide bases, and orientation (*see, Hyrup, et al., 1996. supra*). The synthesis of PNA-DNA chimeras can be performed as described in Hyrup, *et al., 1996. supra* and Finn, *et al., 1996. Nucl Acids Res* 24: 3357-3363. For example, a DNA chain can be synthesized on a solid support using standard phosphoramidite coupling chemistry, and modified nucleoside analogs, *e.g.*, 5'-(4-methoxytrityl)amino-5'-deoxy-thymidine phosphoramidite, can be used between the PNA and the 5' end of DNA. *See, e.g., Mag, et al., 1989. Nucl Acid Res* 17: 5973-5988. PNA monomers are then coupled in a stepwise manner to produce a chimeric molecule with a 5' PNA segment and a 3' DNA segment. *See, e.g., Finn, et al., 1996. supra*. Alternatively, chimeric molecules can be synthesized with a 5' DNA segment and a 3' PNA segment. *See, e.g., Petersen, et al., 1975. Bioorg. Med. Chem. Lett.* 5: 1119-11124.

In other embodiments, the oligonucleotide may include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (*see, e.g., Letsinger, et al., 1989. Proc. Natl. Acad. Sci. U.S.A.* 86: 6553-6556; Lemaitre, *et al., 1987. Proc. Natl. Acad. Sci.* 84: 648-652; PCT Publication No. WO88/09810) or the blood-brain barrier (*see, e.g., PCT Publication No. WO 89/10134*). In addition, oligonucleotides can be modified with hybridization triggered cleavage agents (*see, e.g., Krol, et al., 1988. BioTechniques* 6:958-976) or intercalating agents (*see, e.g., Zon, 1988. Pharm. Res.* 5: 539-549). To this end, the oligonucleotide may be conjugated to another molecule, *e.g.*, a peptide, a hybridization triggered cross-linking agent, a transport agent, a hybridization-triggered cleavage agent, and the like.

NOVX Polypeptides

A polypeptide according to the invention includes a polypeptide including the amino acid sequence of NOVX polypeptides whose sequences are provided in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172. The invention also includes a mutant or variant protein any of whose residues may be changed from the corresponding residues shown in any one of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172, while still encoding a protein that maintains its NOVX activities and physiological functions, or a functional fragment thereof.

In general, a NOVX variant that preserves NOVX-like function includes any variant in which residues at a particular position in the sequence have been substituted by other amino acids, and further include the possibility of inserting an additional residue or residues between two residues of the parent protein as well as the possibility of deleting one or more residues from the parent sequence. Any amino acid substitution, insertion, or deletion is encompassed by the invention. In favorable circumstances, the substitution is a conservative substitution as defined above.

One aspect of the invention pertains to isolated NOVX proteins, and biologically-active portions thereof, or derivatives, fragments, analogs or homologs thereof. Also provided are polypeptide fragments suitable for use as immunogens to raise anti-NOVX antibodies. In one embodiment, native NOVX proteins can be isolated from cells or tissue sources by an appropriate purification scheme using standard protein purification techniques. In another embodiment, NOVX proteins are produced by recombinant DNA techniques. Alternative to recombinant expression, a NOVX protein or polypeptide can be synthesized chemically using standard peptide synthesis techniques.

An "isolated" or "purified" polypeptide or protein or biologically-active portion thereof is substantially free of cellular material or other contaminating proteins from the cell or tissue source from which the NOVX protein is derived, or substantially free from chemical precursors or other chemicals when chemically synthesized. The language "substantially free of cellular material" includes preparations of NOVX proteins in which the protein is separated from cellular components of the cells from which it is isolated or recombinantly-produced. In one embodiment, the language "substantially free of cellular material" includes preparations of NOVX proteins having less than about 30% (by dry weight) of non-NOVX proteins (also referred to herein as a "contaminating protein"), more preferably less than about 20% of non-NOVX proteins, still more preferably less than about

10% of non-NOVX proteins, and most preferably less than about 5% of non-NOVX proteins. When the NOVX protein or biologically-active portion thereof is recombinantly-produced, it is also preferably substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, more preferably less than about 10%, and most preferably less than about 5% of the volume of the NOVX protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins in which the protein is separated from chemical precursors or other chemicals that are involved in the synthesis of the protein. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of NOVX proteins having less than about 30% (by dry weight) of chemical precursors or non-NOVX chemicals, more preferably less than about 20% chemical precursors or non-NOVX chemicals, still more preferably less than about 10% chemical precursors or non-NOVX chemicals, and most preferably less than about 5% chemical precursors or non-NOVX chemicals.

Biologically-active portions of NOVX proteins include peptides comprising amino acid sequences sufficiently homologous to or derived from the amino acid sequences of the NOVX proteins (*e.g.*, the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172) that include fewer amino acids than the full-length NOVX proteins, and exhibit at least one activity of a NOVX protein. Typically, biologically-active portions comprise a domain or motif with at least one activity of the NOVX protein. A biologically-active portion of a NOVX protein can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acid residues in length.

Moreover, other biologically-active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the functional activities of a native NOVX protein.

In an embodiment, the NOVX protein has an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172. In other embodiments, the NOVX protein is substantially homologous to SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172, and retains the functional activity of the protein of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172, yet differs in amino acid sequence due to natural allelic variation or mutagenesis, as described in detail, below. Accordingly, in another embodiment, the NOVX protein is a protein that comprises an amino acid sequence at least about 45% homologous to the amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172, and

retains the functional activity of the NOVX proteins of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172.

Determining Homology Between Two or More Sequences

To determine the percent homology of two amino acid sequences or of two nucleic acids, the sequences are aligned for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are homologous at that position (*i.e.*, as used herein amino acid or nucleic acid "homology" is equivalent to amino acid or nucleic acid "identity").

The nucleic acid sequence homology may be determined as the degree of identity between two sequences. The homology may be determined using computer programs known in the art, such as GAP software provided in the GCG program package. *See*, Needleman and Wunsch, 1970. *J Mol Biol* 48: 443-453. Using GCG GAP software with the following settings for nucleic acid sequence comparison: GAP creation penalty of 5.0 and GAP extension penalty of 0.3, the coding region of the analogous nucleic acid sequences referred to above exhibits a degree of identity preferably of at least 70%, 75%, 80%, 85%, 90%, 95%, 98%, or 99%, with the CDS (encoding) part of the DNA sequence of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172.

The term "sequence identity" refers to the degree to which two polynucleotide or polypeptide sequences are identical on a residue-by-residue basis over a particular region of comparison. The term "percentage of sequence identity" is calculated by comparing two optimally aligned sequences over that region of comparison, determining the number of positions at which the identical nucleic acid base (*e.g.*, A, T, C, G, U, or I, in the case of nucleic acids) occurs in both sequences to yield the number of matched positions, dividing the number of matched positions by the total number of positions in the region of comparison (*i.e.*, the window size), and multiplying the result by 100 to yield the percentage of sequence identity. The term "substantial identity" as used herein denotes a characteristic of a polynucleotide sequence, wherein the polynucleotide comprises a sequence that has at least 80 percent sequence identity, preferably at least 85 percent identity and often 90 to 95 percent

sequence identity, more usually at least 99 percent sequence identity as compared to a reference sequence over a comparison region.

Chimeric and Fusion Proteins

The invention also provides NOVX chimeric or fusion proteins. As used herein, a
5 NOVX "chimeric protein" or "fusion protein" comprises a NOVX polypeptide
operatively-linked to a non-NOVX polypeptide. An "NOVX polypeptide" refers to a
polypeptide having an amino acid sequence corresponding to a NOVX protein of SEQ ID
NO:2*n*, wherein *n* is an integer between 1 and 172, whereas a "non-NOVX polypeptide"
refers to a polypeptide having an amino acid sequence corresponding to a protein that is not
10 substantially homologous to the NOVX protein, *e.g.*, a protein that is different from the
NOVX protein and that is derived from the same or a different organism. Within a NOVX
fusion protein the NOVX polypeptide can correspond to all or a portion of a NOVX protein.
In one embodiment, a NOVX fusion protein comprises at least one biologically-active
portion of a NOVX protein. In another embodiment, a NOVX fusion protein comprises at
15 least two biologically-active portions of a NOVX protein. In yet another embodiment, a
NOVX fusion protein comprises at least three biologically-active portions of a NOVX
protein. Within the fusion protein, the term "operatively-linked" is intended to indicate that
the NOVX polypeptide and the non-NOVX polypeptide are fused in-frame with one another.
The non-NOVX polypeptide can be fused to the N-terminus or C-terminus of the NOVX
20 polypeptide.

In one embodiment, the fusion protein is a GST-NOVX fusion protein in which the
NOVX sequences are fused to the C-terminus of the GST (glutathione S-transferase)
sequences. Such fusion proteins can facilitate the purification of recombinant NOVX
polypeptides.

25 In another embodiment, the fusion protein is a NOVX protein containing a
heterologous signal sequence at its N-terminus. In certain host cells (*e.g.*, mammalian host
cells), expression and/or secretion of NOVX can be increased through use of a heterologous
signal sequence.

In yet another embodiment, the fusion protein is a NOVX-immunoglobulin fusion
30 protein in which the NOVX sequences are fused to sequences derived from a member of the
immunoglobulin protein family. The NOVX-immunoglobulin fusion proteins of the
invention can be incorporated into pharmaceutical compositions and administered to a subject

to inhibit an interaction between a NOVX ligand and a NOVX protein on the surface of a cell, to thereby suppress NOVX-mediated signal transduction *in vivo*. The NOVX-immunoglobulin fusion proteins can be used to affect the bioavailability of a NOVX cognate ligand. Inhibition of the NOVX ligand/NOVX interaction may be useful therapeutically for both the treatment of proliferative and differentiative disorders, as well as modulating (*e.g.* promoting or inhibiting) cell survival. Moreover, the NOVX-immunoglobulin fusion proteins of the invention can be used as immunogens to produce anti-NOVX antibodies in a subject, to purify NOVX ligands, and in screening assays to identify molecules that inhibit the interaction of NOVX with a NOVX ligand.

A NOVX chimeric or fusion protein of the invention can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques, *e.g.*, by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers that give rise to complementary overhangs between two consecutive gene fragments that can subsequently be annealed and reamplified to generate a chimeric gene sequence (*see, e.g.*, Ausubel, *et al.* (eds.) CURRENT PROTOCOLS IN MOLECULAR BIOLOGY, John Wiley & Sons, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST polypeptide). A NOVX-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NOVX protein.

NOVX Agonists and Antagonists

The invention also pertains to variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists. Variants of the NOVX protein can be generated by mutagenesis (*e.g.*, discrete point mutation or truncation of the NOVX protein). An agonist of the NOVX protein can retain substantially the same, or a subset of, the biological activities of the naturally occurring form of the NOVX protein. An antagonist of the NOVX protein can inhibit one or more of the activities of the naturally occurring form of the NOVX protein by, for example, competitively binding to a downstream or upstream

member of a cellular signaling cascade which includes the NOVX protein. Thus, specific biological effects can be elicited by treatment with a variant of limited function. In one embodiment, treatment of a subject with a variant having a subset of the biological activities of the naturally occurring form of the protein has fewer side effects in a subject relative to
5 treatment with the naturally occurring form of the NOVX proteins.

Variants of the NOVX proteins that function as either NOVX agonists (*i.e.*, mimetics) or as NOVX antagonists can be identified by screening combinatorial libraries of mutants (*e.g.*, truncation mutants) of the NOVX proteins for NOVX protein agonist or antagonist activity. In one embodiment, a variegated library of NOVX variants is generated by
10 combinatorial mutagenesis at the nucleic acid level and is encoded by a variegated gene library. A variegated library of NOVX variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential NOVX sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (*e.g.*, for phage display) containing the set of
15 NOVX sequences therein. There are a variety of methods which can be used to produce libraries of potential NOVX variants from a degenerate oligonucleotide sequence. Chemical synthesis of a degenerate gene sequence can be performed in an automatic DNA synthesizer, and the synthetic gene then ligated into an appropriate expression vector. Use of a degenerate set of genes allows for the provision, in one mixture, of all of the sequences encoding the
20 desired set of potential NOVX sequences. Methods for synthesizing degenerate oligonucleotides are well-known within the art. *See, e.g.*, Narang, 1983. *Tetrahedron* 39: 3; Itakura, *et al.*, 1984. *Annu. Rev. Biochem.* 53: 323; Itakura, *et al.*, 1984. *Science* 198: 1056; Ike, *et al.*, 1983. *Nucl. Acids Res.* 11: 477.

Polypeptide Libraries

25 In addition, libraries of fragments of the NOVX protein coding sequences can be used to generate a variegated population of NOVX fragments for screening and subsequent selection of variants of a NOVX protein. In one embodiment, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of a NOVX coding sequence with a nuclease under conditions wherein nicking occurs only about once per
30 molecule, denaturing the double stranded DNA, renaturing the DNA to form double-stranded DNA that can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S_1 nuclease, and ligating the

resulting fragment library into an expression vector. By this method, expression libraries can be derived which encodes N-terminal and internal fragments of various sizes of the NOVX proteins.

Various techniques are known in the art for screening gene products of combinatorial libraries made by point mutations or truncation, and for screening cDNA libraries for gene products having a selected property. Such techniques are adaptable for rapid screening of the gene libraries generated by the combinatorial mutagenesis of NOVX proteins. The most widely used techniques, which are amenable to high throughput analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a new technique that enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify NOVX variants. *See, e.g.,* Arkin and Yourvan, 1992. *Proc. Natl. Acad. Sci. USA* 89: 7811-7815; Delgrave, *et al.*, 1993. *Protein Engineering* 6:327-331.

Anti-NOVX Antibodies

Included in the invention are antibodies to NOVX proteins, or fragments of NOVX proteins. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin (Ig) molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds (immunoreacts with) an antigen. Such antibodies include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, F_{ab} , F_{ab}' and $F_{(ab)2}$ fragments, and an F_{ab} expression library. In general, antibody molecules obtained from humans relates to any of the classes IgG, IgM, IgA, IgE and IgD, which differ from one another by the nature of the heavy chain present in the molecule. Certain classes have subclasses as well, such as IgG₁, IgG₂, and others. Furthermore, in humans, the light chain may be a kappa chain or a lambda chain. Reference herein to antibodies includes a reference to all such classes, subclasses and types of human antibody species.

An isolated protein of the invention intended to serve as an antigen, or a portion or fragment thereof, can be used as an immunogen to generate antibodies that immunospecifically bind the antigen, using standard techniques for polyclonal and monoclonal antibody preparation. The full-length protein can be used or, alternatively, the

invention provides antigenic peptide fragments of the antigen for use as immunogens. An antigenic peptide fragment comprises at least 6 amino acid residues of the amino acid sequence of the full length protein, such as an amino acid sequence of SEQ ID NO:2*n*, wherein *n* is an integer between 1 and 172, and encompasses an epitope thereof such that an antibody raised against the peptide forms a specific immune complex with the full length protein or with any fragment that contains the epitope. Preferably, the antigenic peptide comprises at least 10 amino acid residues, or at least 15 amino acid residues, or at least 20 amino acid residues, or at least 30 amino acid residues. Preferred epitopes encompassed by the antigenic peptide are regions of the protein that are located on its surface; commonly these are hydrophilic regions.

In certain embodiments of the invention, at least one epitope encompassed by the antigenic peptide is a region of NOVX that is located on the surface of the protein, *e.g.*, a hydrophilic region. A hydrophobicity analysis of the human NOVX protein sequence will indicate which regions of a NOVX polypeptide are particularly hydrophilic and, therefore, are likely to encode surface residues useful for targeting antibody production. As a means for targeting antibody production, hydropathy plots showing regions of hydrophilicity and hydrophobicity may be generated by any method well known in the art, including, for example, the Kyte Doolittle or the Hopp Woods methods, either with or without Fourier transformation. See, *e.g.*, Hopp and Woods, 1981, *Proc. Nat. Acad. Sci. USA* 78: 3824-3828; Kyte and Doolittle 1982, *J. Mol. Biol.* 157: 105-142, each incorporated herein by reference in their entirety. Antibodies that are specific for one or more domains within an antigenic protein, or derivatives, fragments, analogs or homologs thereof, are also provided herein.

The term "epitope" includes any protein determinant capable of specific binding to an immunoglobulin or T-cell receptor. Epitopic determinants usually consist of chemically active surface groupings of molecules such as amino acids or sugar side chains and usually have specific three dimensional structural characteristics, as well as specific charge characteristics. A NOVX polypeptide or a fragment thereof comprises at least one antigenic epitope. An anti-NOVX antibody of the present invention is said to specifically bind to antigen NOVX when the equilibrium binding constant (K_D) is $\leq 1 \mu\text{M}$, preferably $\leq 100 \text{ nM}$, more preferably $\leq 10 \text{ nM}$, and most preferably $\leq 100 \text{ pM}$ to about 1 pM , as measured by assays such as radioligand binding assays or similar assays known to those skilled in the art.

A protein of the invention, or a derivative, fragment, analog, homolog or ortholog thereof, may be utilized as an immunogen in the generation of antibodies that immunospecifically bind these protein components.

Various procedures known within the art may be used for the production of polyclonal or monoclonal antibodies directed against a protein of the invention, or against derivatives, fragments, analogs homologs or orthologs thereof (see, for example, *Antibodies: A Laboratory Manual*, Harlow E, and Lane D, 1988, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, incorporated herein by reference). Some of these antibodies are discussed below.

10 **Polyclonal Antibodies**

For the production of polyclonal antibodies, various suitable host animals (*e.g.*, rabbit, goat, mouse or other mammal) may be immunized by one or more injections with the native protein, a synthetic variant thereof, or a derivative of the foregoing. An appropriate immunogenic preparation can contain, for example, the naturally occurring immunogenic protein, a chemically synthesized polypeptide representing the immunogenic protein, or a recombinantly expressed immunogenic protein. Furthermore, the protein may be conjugated to a second protein known to be immunogenic in the mammal being immunized. Examples of such immunogenic proteins include but are not limited to keyhole limpet hemocyanin, serum albumin, bovine thyroglobulin, and soybean trypsin inhibitor. The preparation can further include an adjuvant. Various adjuvants used to increase the immunological response include, but are not limited to, Freund's (complete and incomplete), mineral gels (*e.g.*, aluminum hydroxide), surface active substances (*e.g.*, lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, dinitrophenol, *etc.*), adjuvants usable in humans such as Bacille Calmette-Guerin and *Corynebacterium parvum*, or similar immunostimulatory agents. Additional examples of adjuvants which can be employed include MPL-TDM adjuvant (monophosphoryl Lipid A, synthetic trehalose dicorynomycolate).

The polyclonal antibody molecules directed against the immunogenic protein can be isolated from the mammal (*e.g.*, from the blood) and further purified by well known techniques, such as affinity chromatography using protein A or protein G, which provide primarily the IgG fraction of immune serum. Subsequently, or alternatively, the specific antigen which is the target of the immunoglobulin sought, or an epitope thereof, may be immobilized on a column to purify the immune specific antibody by immunoaffinity

chromatography. Purification of immunoglobulins is discussed, for example, by D. Wilkinson (The Scientist, published by The Scientist, Inc., Philadelphia PA, Vol. 14, No. 8 (April 17, 2000), pp. 25-28).

Monoclonal Antibodies

5 The term "monoclonal antibody" (MAb) or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one molecular species of antibody molecule consisting of a unique light chain gene product and a unique heavy chain gene product. In particular, the complementarity determining regions (CDRs) of the monoclonal antibody are identical in all the molecules of the population. MABs thus
10 contain an antigen binding site capable of immunoreacting with a particular epitope of the antigen characterized by a unique binding affinity for it.

Monoclonal antibodies can be prepared using hybridoma methods, such as those described by Kohler and Milstein, Nature, 256:495 (1975). In a hybridoma method, a mouse, hamster, or other appropriate host animal, is typically immunized with an immunizing agent
15 to elicit lymphocytes that produce or are capable of producing antibodies that will specifically bind to the immunizing agent. Alternatively, the lymphocytes can be immunized in vitro.

The immunizing agent will typically include the protein antigen, a fragment thereof or a fusion protein thereof. Generally, either peripheral blood lymphocytes are used if cells
20 of human origin are desired, or spleen cells or lymph node cells are used if non-human mammalian sources are desired. The lymphocytes are then fused with an immortalized cell line using a suitable fusing agent, such as polyethylene glycol, to form a hybridoma cell (Goding, Monoclonal Antibodies: Principles and Practice, Academic Press, (1986) pp. 59-103). Immortalized cell lines are usually transformed mammalian cells, particularly
25 myeloma cells of rodent, bovine and human origin. Usually, rat or mouse myeloma cell lines are employed. The hybridoma cells can be cultured in a suitable culture medium that preferably contains one or more substances that inhibit the growth or survival of the unfused, immortalized cells. For example, if the parental cells lack the enzyme hypoxanthine guanine phosphoribosyl transferase (HGPRT or HPRT), the culture medium for the hybridomas
30 typically will include hypoxanthine, aminopterin, and thymidine ("HAT medium"), which substances prevent the growth of HGPRT-deficient cells.

Preferred immortalized cell lines are those that fuse efficiently, support stable high level expression of antibody by the selected antibody-producing cells, and are sensitive to a medium such as HAT medium. More preferred immortalized cell lines are murine myeloma lines, which can be obtained, for instance, from the Salk Institute Cell Distribution Center,
5 San Diego, California and the American Type Culture Collection, Manassas, Virginia. Human myeloma and mouse-human heteromyeloma cell lines also have been described for the production of human monoclonal antibodies (Kozbor, J. Immunol., 133:3001 (1984); Brodeur et al., Monoclonal Antibody Production Techniques and Applications, Marcel Dekker, Inc., New York, (1987) pp. 51-63).

10 The culture medium in which the hybridoma cells are cultured can then be assayed for the presence of monoclonal antibodies directed against the antigen. Preferably, the binding specificity of monoclonal antibodies produced by the hybridoma cells is determined by immunoprecipitation or by an in vitro binding assay, such as radioimmunoassay (RIA) or enzyme-linked immunoabsorbent assay (ELISA). Such techniques and assays are known in
15 the art. The binding affinity of the monoclonal antibody can, for example, be determined by the Scatchard analysis of Munson and Pollard, Anal. Biochem., 107:220 (1980). It is an objective, especially important in therapeutic applications of monoclonal antibodies, to identify antibodies having a high degree of specificity and a high binding affinity for the target antigen.

20 After the desired hybridoma cells are identified, the clones can be subcloned by limiting dilution procedures and grown by standard methods (Goding, 1986). Suitable culture media for this purpose include, for example, Dulbecco's Modified Eagle's Medium and RPMI-1640 medium. Alternatively, the hybridoma cells can be grown in vivo as ascites in a mammal.

25 The monoclonal antibodies secreted by the subclones can be isolated or purified from the culture medium or ascites fluid by conventional immunoglobulin purification procedures such as, for example, protein A-Sepharose, hydroxylapatite chromatography, gel electrophoresis, dialysis, or affinity chromatography.

The monoclonal antibodies can also be made by recombinant DNA methods, such as
30 those described in U.S. Patent No. 4,816,567. DNA encoding the monoclonal antibodies of the invention can be readily isolated and sequenced using conventional procedures (e.g., by using oligonucleotide probes that are capable of binding specifically to genes encoding the heavy and light chains of murine antibodies). The hybridoma cells of the invention serve as a

preferred source of such DNA. Once isolated, the DNA can be placed into expression vectors, which are then transfected into host cells such as simian COS cells, Chinese hamster ovary (CHO) cells, or myeloma cells that do not otherwise produce immunoglobulin protein, to obtain the synthesis of monoclonal antibodies in the recombinant host cells. The DNA
5 also can be modified, for example, by substituting the coding sequence for human heavy and light chain constant domains in place of the homologous murine sequences (U.S. Patent No. 4,816,567; Morrison, Nature 368, 812-13 (1994)) or by covalently joining to the immunoglobulin coding sequence all or part of the coding sequence for a non-immunoglobulin polypeptide. Such a non-immunoglobulin polypeptide can be
10 substituted for the constant domains of an antibody of the invention, or can be substituted for the variable domains of one antigen-combining site of an antibody of the invention to create a chimeric bivalent antibody.

Humanized Antibodies

The antibodies directed against the protein antigens of the invention can further
15 comprise humanized antibodies or human antibodies. These antibodies are suitable for administration to humans without engendering an immune response by the human against the administered immunoglobulin. Humanized forms of antibodies are chimeric immunoglobulins, immunoglobulin chains or fragments thereof (such as Fv, Fab, Fab', F(ab')₂ or other antigen-binding subsequences of antibodies) that are principally comprised of the
20 sequence of a human immunoglobulin, and contain minimal sequence derived from a non-human immunoglobulin. Humanization can be performed following the method of Winter and co-workers (Jones et al., Nature, 321:522-525 (1986); Riechmann et al., Nature, 332:323-327 (1988); Verhoeven et al., Science, 239:1534-1536 (1988)), by substituting rodent CDRs or CDR sequences for the corresponding sequences of a human antibody. (See
25 also U.S. Patent No. 5,225,539.) In some instances, Fv framework residues of the human immunoglobulin are replaced by corresponding non-human residues. Humanized antibodies can also comprise residues which are found neither in the recipient antibody nor in the imported CDR or framework sequences. In general, the humanized antibody will comprise substantially all of at least one, and typically two, variable domains, in which all or
30 substantially all of the CDR regions correspond to those of a non-human immunoglobulin and all or substantially all of the framework regions are those of a human immunoglobulin consensus sequence. The humanized antibody optimally also will comprise at least a portion

of an immunoglobulin constant region (Fc), typically that of a human immunoglobulin (Jones et al., 1986; Riechmann et al., 1988; and Presta, Curr. Op. Struct. Biol., 2:593-596 (1992)).

Human Antibodies

Fully human antibodies essentially relate to antibody molecules in which the entire
5 sequence of both the light chain and the heavy chain, including the CDRs, arise from human
genes. Such antibodies are termed "human antibodies", or "fully human antibodies" herein.
Human monoclonal antibodies can be prepared by the trioma technique; the human B-cell
hybridoma technique (see Kozbor, et al., 1983 Immunol Today 4: 72) and the EBV
hybridoma technique to produce human monoclonal antibodies (see Cole, et al., 1985 In:
10 MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp. 77-96). Human
monoclonal antibodies may be utilized in the practice of the present invention and may be
produced by using human hybridomas (see Cote, et al., 1983. Proc Natl Acad Sci USA 80:
2026-2030) or by transforming human B-cells with Epstein Barr Virus in vitro (see Cole, et
al., 1985 In: MONOCLONAL ANTIBODIES AND CANCER THERAPY, Alan R. Liss, Inc., pp.
15 77-96).

In addition, human antibodies can also be produced using additional techniques,
including phage display libraries (Hoogenboom and Winter, J. Mol. Biol., 227:381 (1991);
Marks et al., J. Mol. Biol., 222:581 (1991)). Similarly, human antibodies can be made by
introducing human immunoglobulin loci into transgenic animals, *e.g.*, mice in which the
20 endogenous immunoglobulin genes have been partially or completely inactivated. Upon
challenge, human antibody production is observed, which closely resembles that seen in
humans in all respects, including gene rearrangement, assembly, and antibody repertoire.
This approach is described, for example, in U.S. Patent Nos. 5,545,807; 5,545,806;
5,569,825; 5,625,126; 5,633,425; 5,661,016, and in Marks et al. (Bio/Technology 10,
25 779-783 (1992)); Lonberg et al. (Nature 368 856-859 (1994)); Morrison (Nature 368,
812-13 (1994)); Fishwild et al. (Nature Biotechnology 14, 845-51 (1996)); Neuberger
(Nature Biotechnology 14, 826 (1996)); and Lonberg and Huszar (Intern. Rev. Immunol. 13
65-93 (1995)).

Human antibodies may additionally be produced using transgenic nonhuman animals
30 which are modified so as to produce fully human antibodies rather than the animal's
endogenous antibodies in response to challenge by an antigen. (See PCT publication
WO94/02602). The endogenous genes encoding the heavy and light immunoglobulin chains

in the nonhuman host have been incapacitated, and active loci encoding human heavy and light chain immunoglobulins are inserted into the host's genome. The human genes are incorporated, for example, using yeast artificial chromosomes containing the requisite human DNA segments. An animal which provides all the desired modifications is then obtained as progeny by crossbreeding intermediate transgenic animals containing fewer than the full complement of the modifications. The preferred embodiment of such a nonhuman animal is a mouse, and is termed the XenomouseTM as disclosed in PCT publications WO 96/33735 and WO 96/34096. This animal produces B cells which secrete fully human immunoglobulins. The antibodies can be obtained directly from the animal after immunization with an immunogen of interest, as, for example, a preparation of a polyclonal antibody, or alternatively from immortalized B cells derived from the animal, such as hybridomas producing monoclonal antibodies. Additionally, the genes encoding the immunoglobulins with human variable regions can be recovered and expressed to obtain the antibodies directly, or can be further modified to obtain analogs of antibodies such as, for example, single chain Fv molecules.

An example of a method of producing a nonhuman host, exemplified as a mouse, lacking expression of an endogenous immunoglobulin heavy chain is disclosed in U.S. Patent No. 5,939,598. It can be obtained by a method including deleting the J segment genes from at least one endogenous heavy chain locus in an embryonic stem cell to prevent rearrangement of the locus and to prevent formation of a transcript of a rearranged immunoglobulin heavy chain locus, the deletion being effected by a targeting vector containing a gene encoding a selectable marker; and producing from the embryonic stem cell a transgenic mouse whose somatic and germ cells contain the gene encoding the selectable marker.

A method for producing an antibody of interest, such as a human antibody, is disclosed in U.S. Patent No. 5,916,771. It includes introducing an expression vector that contains a nucleotide sequence encoding a heavy chain into one mammalian host cell in culture, introducing an expression vector containing a nucleotide sequence encoding a light chain into another mammalian host cell, and fusing the two cells to form a hybrid cell. The hybrid cell expresses an antibody containing the heavy chain and the light chain.

In a further improvement on this procedure, a method for identifying a clinically relevant epitope on an immunogen, and a correlative method for selecting an antibody that

binds immunospecifically to the relevant epitope with high affinity, are disclosed in PCT publication WO 99/53049.

F_{ab} Fragments and Single Chain Antibodies

According to the invention, techniques can be adapted for the production of single-chain antibodies specific to an antigenic protein of the invention (see *e.g.*, U.S. Patent No. 4,946,778). In addition, methods can be adapted for the construction of F_{ab} expression libraries (see *e.g.*, Huse, et al., 1989 Science 246: 1275-1281) to allow rapid and effective identification of monoclonal F_{ab} fragments with the desired specificity for a protein or derivatives, fragments, analogs or homologs thereof. Antibody fragments that contain the idiotype to a protein antigen may be produced by techniques known in the art including, but not limited to: (i) an F_{(ab)₂} fragment produced by pepsin digestion of an antibody molecule; (ii) an F_{ab} fragment generated by reducing the disulfide bridges of an F_{(ab)₂} fragment; (iii) an F_{ab} fragment generated by the treatment of the antibody molecule with papain and a reducing agent and (iv) F_v fragments.

Bispecific Antibodies

Bispecific antibodies are monoclonal, preferably human or humanized, antibodies that have binding specificities for at least two different antigens. In the present case, one of the binding specificities is for an antigenic protein of the invention. The second binding target is any other antigen, and advantageously is a cell-surface protein or receptor or receptor subunit.

Methods for making bispecific antibodies are known in the art. Traditionally, the recombinant production of bispecific antibodies is based on the co-expression of two immunoglobulin heavy-chain/light-chain pairs, where the two heavy chains have different specificities (Milstein and Cuello, Nature, 305:537-539 (1983)). Because of the random assortment of immunoglobulin heavy and light chains, these hybridomas (quadromas) produce a potential mixture of ten different antibody molecules, of which only one has the correct bispecific structure. The purification of the correct molecule is usually accomplished by affinity chromatography steps. Similar procedures are disclosed in WO 93/08829, published 13 May 1993, and in Traunecker et al., EMBO J., 10:3655-3659 (1991).

Antibody variable domains with the desired binding specificities (antibody-antigen combining sites) can be fused to immunoglobulin constant domain sequences. The fusion preferably is with an immunoglobulin heavy-chain constant domain, comprising at least part

of the hinge, CH2, and CH3 regions. It is preferred to have the first heavy-chain constant region (CH1) containing the site necessary for light-chain binding present in at least one of the fusions. DNAs encoding the immunoglobulin heavy-chain fusions and, if desired, the immunoglobulin light chain, are inserted into separate expression vectors, and are
5 co-transfected into a suitable host organism. For further details of generating bispecific antibodies see, for example, Suresh et al., *Methods in Enzymology*, 121:210 (1986).

According to another approach described in WO 96/27011, the interface between a pair of antibody molecules can be engineered to maximize the percentage of heterodimers which are recovered from recombinant cell culture. The preferred interface comprises at least
10 a part of the CH3 region of an antibody constant domain. In this method, one or more small amino acid side chains from the interface of the first antibody molecule are replaced with larger side chains (*e.g.* tyrosine or tryptophan). Compensatory "cavities" of identical or similar size to the large side chain(s) are created on the interface of the second antibody molecule by replacing large amino acid side chains with smaller ones (*e.g.* alanine or
15 threonine). This provides a mechanism for increasing the yield of the heterodimer over other unwanted end-products such as homodimers.

Bispecific antibodies can be prepared as full length antibodies or antibody fragments (*e.g.* F(ab')₂ bispecific antibodies). Techniques for generating bispecific antibodies from antibody fragments have been described in the literature. For example, bispecific antibodies
20 can be prepared using chemical linkage. Brennan et al., *Science* 229:81 (1985) describe a procedure wherein intact antibodies are proteolytically cleaved to generate F(ab')₂ fragments. These fragments are reduced in the presence of the dithiol complexing agent sodium arsenite to stabilize vicinal dithiols and prevent intermolecular disulfide formation. The Fab' fragments generated are then converted to thionitrobenzoate (TNB) derivatives. One of the
25 Fab'-TNB derivatives is then reconverted to the Fab'-thiol by reduction with mercaptoethylamine and is mixed with an equimolar amount of the other Fab'-TNB derivative to form the bispecific antibody. The bispecific antibodies produced can be used as agents for the selective immobilization of enzymes.

Additionally, Fab' fragments can be directly recovered from *E. coli* and chemically
30 coupled to form bispecific antibodies. Shalaby et al., *J. Exp. Med.* 175:217-225 (1992) describe the production of a fully humanized bispecific antibody F(ab')₂ molecule. Each Fab' fragment was separately secreted from *E. coli* and subjected to directed chemical coupling in vitro to form the bispecific antibody. The bispecific antibody thus formed was

able to bind to cells overexpressing the ErbB2 receptor and normal human T cells, as well as trigger the lytic activity of human cytotoxic lymphocytes against human breast tumor targets.

Various techniques for making and isolating bispecific antibody fragments directly from recombinant cell culture have also been described. For example, bispecific antibodies
5 have been produced using leucine zippers. Kostelny et al., J. Immunol. 148(5):1547-1553 (1992). The leucine zipper peptides from the Fos and Jun proteins were linked to the Fab' portions of two different antibodies by gene fusion. The antibody homodimers were reduced at the hinge region to form monomers and then re-oxidized to form the antibody heterodimers. This method can also be utilized for the production of antibody homodimers.
10 The "diabody" technology described by Hollinger et al., Proc. Natl. Acad. Sci. USA 90:6444-6448 (1993) has provided an alternative mechanism for making bispecific antibody fragments. The fragments comprise a heavy-chain variable domain (V_H) connected to a light-chain variable domain (V_L) by a linker which is too short to allow pairing between the two domains on the same chain. Accordingly, the V_H and V_L domains of one fragment are
15 forced to pair with the complementary V_L and V_H domains of another fragment, thereby forming two antigen-binding sites. Another strategy for making bispecific antibody fragments by the use of single-chain Fv (sFv) dimers has also been reported. See, Gruber et al., J. Immunol. 152:5368 (1994).

Antibodies with more than two valencies are contemplated. For example, trispecific
20 antibodies can be prepared. Tutt et al., J. Immunol. 147:60 (1991).

Exemplary bispecific antibodies can bind to two different epitopes, at least one of which originates in the protein antigen of the invention. Alternatively, an anti-antigenic arm of an immunoglobulin molecule can be combined with an arm which binds to a triggering molecule on a leukocyte such as a T-cell receptor molecule (*e.g.* CD2, CD3, CD28, or B7), or
25 Fc receptors for IgG (FcγR), such as FcγRI (CD64), FcγRII (CD32) and FcγRIII (CD16) so as to focus cellular defense mechanisms to the cell expressing the particular antigen.

Bispecific antibodies can also be used to direct cytotoxic agents to cells which express a particular antigen. These antibodies possess an antigen-binding arm and an arm which binds a cytotoxic agent or a radionuclide chelator, such as EOTUBE, DPTA, DOTA, or TETA.
30 Another bispecific antibody of interest binds the protein antigen described herein and further binds tissue factor (TF).

Heteroconjugate Antibodies

Heteroconjugate antibodies are also within the scope of the present invention. Heteroconjugate antibodies are composed of two covalently joined antibodies. Such antibodies have, for example, been proposed to target immune system cells to unwanted cells (U.S. Patent No. 4,676,980), and for treatment of HIV infection (WO 91/00360; WO 5 92/200373; EP 03089). It is contemplated that the antibodies can be prepared in vitro using known methods in synthetic protein chemistry, including those involving crosslinking agents. For example, immunotoxins can be constructed using a disulfide exchange reaction or by forming a thioether bond. Examples of suitable reagents for this purpose include 10 iminothiolate and methyl-4-mercaptobutyrimidate and those disclosed, for example, in U.S. Patent No. 4,676,980.

Effector Function Engineering

It can be desirable to modify the antibody of the invention with respect to effector function, so as to enhance, *e.g.*, the effectiveness of the antibody in treating cancer. For 15 example, cysteine residue(s) can be introduced into the Fc region, thereby allowing interchain disulfide bond formation in this region. The homodimeric antibody thus generated can have improved internalization capability and/or increased complement-mediated cell killing and antibody-dependent cellular cytotoxicity (ADCC). See Caron et al., J. Exp Med., 176: 1191-1195 (1992) and Shopes, J. Immunol., 148: 2918-2922 (1992). Homodimeric 20 antibodies with enhanced anti-tumor activity can also be prepared using heterobifunctional cross-linkers as described in Wolff et al. Cancer Research, 53: 2560-2565 (1993). Alternatively, an antibody can be engineered that has dual Fc regions and can thereby have enhanced complement lysis and ADCC capabilities. See Stevenson et al., Anti-Cancer Drug Design, 3: 219-230 (1989).

25 Immunoconjugates

The invention also pertains to immunoconjugates comprising an antibody conjugated to a cytotoxic agent such as a chemotherapeutic agent, toxin (*e.g.*, an enzymatically active toxin of bacterial, fungal, plant, or animal origin, or fragments thereof), or a radioactive isotope (*i.e.*, a radioconjugate).

30 Chemotherapeutic agents useful in the generation of such immunoconjugates have been described above. Enzymatically active toxins and fragments thereof that can be used include diphtheria A chain, nonbinding active fragments of diphtheria toxin, exotoxin A

chain (from *Pseudomonas aeruginosa*), ricin A chain, abrin A chain, modeccin A chain, alpha-sarcin, *Aleurites fordii* proteins, dianthin proteins, *Phytolaca americana* proteins (PAPI, PAPII, and PAP-S), momordica charantia inhibitor, curcin, crotin, *sapaonaria officinalis* inhibitor, gelonin, mitogellin, restrictocin, phenomycin, enomycin, and the tricothecenes. A
5 variety of radionuclides are available for the production of radioconjugated antibodies. Examples include ^{212}Bi , ^{131}I , ^{131}In , ^{90}Y , and ^{186}Re .

Conjugates of the antibody and cytotoxic agent are made using a variety of bifunctional protein-coupling agents such as N-succinimidyl-3-(2-pyridyldithiol) propionate (SPDP), iminothiolane (IT), bifunctional derivatives of imidoesters (such as dimethyl
10 adipimidate HCL), active esters (such as disuccinimidyl suberate), aldehydes (such as glutaraldehyde), bis-azido compounds (such as bis (p-azidobenzoyl) hexanediamine), bis-diazonium derivatives (such as bis-(p-diazoniumbenzoyl)-ethylenediamine), diisocyanates (such as tolyene 2,6-diisocyanate), and bis-active fluorine compounds (such as 1,5-difluoro-2,4-dinitrobenzene). For example, a ricin immunotoxin can be prepared as
15 described in Vitetta et al., Science, 238: 1098 (1987). Carbon-14-labeled 1-isothiocyanatobenzyl-3-methyldiethylene triaminepentaacetic acid (MX-DTPA) is an exemplary chelating agent for conjugation of radionucleotide to the antibody. See WO94/11026.

In another embodiment, the antibody can be conjugated to a "receptor" (such
20 streptavidin) for utilization in tumor pretargeting wherein the antibody-receptor conjugate is administered to the patient, followed by removal of unbound conjugate from the circulation using a clearing agent and then administration of a "ligand" (*e.g.*, avidin) that is in turn conjugated to a cytotoxic agent.

Immunoliposomes

25 The antibodies disclosed herein can also be formulated as immunoliposomes. Liposomes containing the antibody are prepared by methods known in the art, such as described in Epstein et al., *Proc. Natl. Acad. Sci. USA*, 82: 3688 (1985); Hwang et al., *Proc. Natl. Acad. Sci. USA*, 77: 4030 (1980); and U.S. Pat. Nos. 4,485,045 and 4,544,545. Liposomes with enhanced circulation time are disclosed in U.S. Patent No. 5,013,556.

30 Particularly useful liposomes can be generated by the reverse-phase evaporation method with a lipid composition comprising phosphatidylcholine, cholesterol, and PEG-derivatized phosphatidylethanolamine (PEG-PE). Liposomes are extruded through

filters of defined pore size to yield liposomes with the desired diameter. Fab' fragments of the antibody of the present invention can be conjugated to the liposomes as described in Martin et al., J. Biol. Chem., 257: 286-288 (1982) via a disulfide-interchange reaction. A chemotherapeutic agent (such as Doxorubicin) is optionally contained within the liposome.

5 See Gabizon *et al.*, J. National Cancer Inst., 81(19): 1484 (1989).

Diagnostic Applications of Antibodies Directed Against the Proteins of the Invention

In one embodiment, methods for the screening of antibodies that possess the desired specificity include, but are not limited to, enzyme linked immunosorbent assay (ELISA) and
10 other immunologically mediated techniques known within the art. In a specific embodiment, selection of antibodies that are specific to a particular domain of an NOVX protein is facilitated by generation of hybridomas that bind to the fragment of an NOVX protein possessing such a domain. Thus, antibodies that are specific for a desired domain within an NOVX protein, or derivatives, fragments, analogs or homologs thereof, are also provided
15 herein.

Antibodies directed against a NOVX protein of the invention may be used in methods known within the art relating to the localization and/or quantitation of a NOVX protein (*e.g.*, for use in measuring levels of the NOVX protein within appropriate physiological samples, for use in diagnostic methods, for use in imaging the protein, and the like). In a given
20 embodiment, antibodies specific to a NOVX protein, or derivative, fragment, analog or homolog thereof, that contain the antibody derived antigen binding domain, are utilized as pharmacologically active compounds (referred to hereinafter as "Therapeutics").

An antibody specific for a NOVX protein of the invention (*e.g.*, a monoclonal antibody or a polyclonal antibody) can be used to isolate a NOVX polypeptide by standard
25 techniques, such as immunoaffinity, chromatography or immunoprecipitation. An antibody to a NOVX polypeptide can facilitate the purification of a natural NOVX antigen from cells, or of a recombinantly produced NOVX antigen expressed in host cells. Moreover, such an anti-NOVX antibody can be used to detect the antigenic NOVX protein (*e.g.*, in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the
30 antigenic NOVX protein. Antibodies directed against a NOVX protein can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be

facilitated by coupling (*i.e.*, physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, 5 β -galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include 10 luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I , ^{131}I , ^{35}S or ^3H .

Antibody Therapeutics

Antibodies of the invention, including polyclonal, monoclonal, humanized and fully human antibodies, may be used as therapeutic agents. Such agents will generally be employed 15 to treat or prevent a disease or pathology in a subject. An antibody preparation, preferably one having high specificity and high affinity for its target antigen, is administered to the subject and will generally have an effect due to its binding with the target. Such an effect may be one of two kinds, depending on the specific nature of the interaction between the given antibody molecule and the target antigen in question. In the first instance, 20 administration of the antibody may abrogate or inhibit the binding of the target with an endogenous ligand to which it naturally binds. In this case, the antibody binds to the target and masks a binding site of the naturally occurring ligand, wherein the ligand serves as an effector molecule. Thus the receptor mediates a signal transduction pathway for which ligand is responsible.

25 Alternatively, the effect may be one in which the antibody elicits a physiological result by virtue of binding to an effector binding site on the target molecule. In this case the target, a receptor having an endogenous ligand which may be absent or defective in the disease or pathology, binds the antibody as a surrogate effector ligand, initiating a receptor-based signal transduction event by the receptor.

30 A therapeutically effective amount of an antibody of the invention relates generally to the amount needed to achieve a therapeutic objective. As noted above, this may be a binding interaction between the antibody and its target antigen that, in certain cases, interferes with

the functioning of the target, and in other cases, promotes a physiological response. The amount required to be administered will furthermore depend on the binding affinity of the antibody for its specific antigen, and will also depend on the rate at which an administered antibody is depleted from the free volume other subject to which it is administered. Common
5 ranges for therapeutically effective dosing of an antibody or antibody fragment of the invention may be, by way of nonlimiting example, from about 0.1 mg/kg body weight to about 50 mg/kg body weight. Common dosing frequencies may range, for example, from twice daily to once a week.

Pharmaceutical Compositions of Antibodies

10 Antibodies specifically binding a protein of the invention, as well as other molecules identified by the screening assays disclosed herein, can be administered for the treatment of various disorders in the form of pharmaceutical compositions. Principles and considerations involved in preparing such compositions, as well as guidance in the choice of components are provided, for example, in Remington : The Science And Practice Of Pharmacy 19th ed.
15 (Alfonso R. Gennaro, et al., editors) Mack Pub. Co., Easton, Pa. : 1995; Drug Absorption Enhancement : Concepts, Possibilities, Limitations, And Trends, Harwood Academic Publishers, Langhorne, Pa., 1994; and Peptide And Protein Drug Delivery (Advances In Parenteral Sciences, Vol. 4), 1991, M. Dekker, New York.

If the antigenic protein is intracellular and whole antibodies are used as inhibitors,
20 internalizing antibodies are preferred. However, liposomes can also be used to deliver the antibody, or an antibody fragment, into cells. Where antibody fragments are used, the smallest inhibitory fragment that specifically binds to the binding domain of the target protein is preferred. For example, based upon the variable-region sequences of an antibody, peptide molecules can be designed that retain the ability to bind the target protein sequence. Such
25 peptides can be synthesized chemically and/or produced by recombinant DNA technology. See, e.g., Marasco et al., Proc. Natl. Acad. Sci. USA, 90: 7889-7893 (1993). The formulation herein can also contain more than one active compound as necessary for the particular indication being treated, preferably those with complementary activities that do not adversely affect each other. Alternatively, or in addition, the composition can comprise an
30 agent that enhances its function, such as, for example, a cytotoxic agent, cytokine, chemotherapeutic agent, or growth-inhibitory agent. Such molecules are suitably present in combination in amounts that are effective for the purpose intended.

The active ingredients can also be entrapped in microcapsules prepared, for example, by coacervation techniques or by interfacial polymerization, for example, hydroxymethylcellulose or gelatin-microcapsules and poly-(methacrylate) microcapsules, respectively, in colloidal drug delivery systems (for example, liposomes, albumin microspheres, microemulsions, nano-particles, and nanocapsules) or in macroemulsions.

The formulations to be used for in vivo administration must be sterile. This is readily accomplished by filtration through sterile filtration membranes.

Sustained-release preparations can be prepared. Suitable examples of sustained-release preparations include semipermeable matrices of solid hydrophobic polymers containing the antibody, which matrices are in the form of shaped articles, e.g., films, or microcapsules. Examples of sustained-release matrices include polyesters, hydrogels (for example, poly(2-hydroxyethyl-methacrylate), or poly(vinylalcohol)), polylactides (U.S. Pat. No. 3,773,919), copolymers of L-glutamic acid and γ ethyl-L-glutamate, non-degradable ethylene-vinyl acetate, degradable lactic acid-glycolic acid copolymers such as the LUPRON DEPOTTM (injectable microspheres composed of lactic acid-glycolic acid copolymer and leuprolide acetate), and poly-D-(-)-3-hydroxybutyric acid. While polymers such as ethylene-vinyl acetate and lactic acid-glycolic acid enable release of molecules for over 100 days, certain hydrogels release proteins for shorter time periods.

ELISA Assay

An agent for detecting an analyte protein is an antibody capable of binding to an analyte protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (e.g., F_{ab} or $F_{(ab)2}$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids

present within a subject. Included within the usage of the term "biological sample", therefore, is blood and a fraction or component of blood including blood serum, blood plasma, or lymph. That is, the detection method of the invention can be used to detect an analyte mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*.

5 For example, *in vitro* techniques for detection of an analyte mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of an analyte protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of an analyte genomic DNA include Southern hybridizations. Procedures for conducting
10 immunoassays are described, for example in "ELISA: Theory and Practice: Methods in Molecular Biology", Vol. 42, J. R. Crowther (Ed.) Human Press, Totowa, NJ, 1995; "Immunoassay", E. Diamandis and T. Christopoulos, Academic Press, Inc., San Diego, CA, 1996; and "Practice and Theory of Enzyme Immunoassays", P. Tijssen, Elsevier Science Publishers, Amsterdam, 1985. Furthermore, *in vivo* techniques for detection of an analyte
15 protein include introducing into a subject a labeled anti-analyte protein antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

NOVX Recombinant Expression Vectors and Host Cells

Another aspect of the invention pertains to vectors, preferably expression vectors,
20 containing a nucleic acid encoding a NOVX protein, or derivatives, fragments, analogs or homologs thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA
25 segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome.
30 Moreover, certain vectors are capable of directing the expression of genes to which they are operatively-linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of

plasmids. In the present specification, "plasmid" and "vector" can be used interchangeably as the plasmid is the most commonly used form of vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a host cell, which means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, that is operatively-linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably-linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner that allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell).

The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Regulatory sequences include those that direct constitutive expression of a nucleotide sequence in many types of host cell and those that direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, *etc.* The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, including fusion proteins or peptides, encoded by nucleic acids as described herein (*e.g.*, NOVX proteins, mutant forms of NOVX proteins, fusion proteins, *etc.*).

The recombinant expression vectors of the invention can be designed for expression of NOVX proteins in prokaryotic or eukaryotic cells. For example, NOVX proteins can be expressed in bacterial cells such as *Escherichia coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990). Alternatively, the recombinant expression vector can be

transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression of proteins in prokaryotes is most often carried out in *Escherichia coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, usually to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: (i) to increase expression of recombinant protein; (ii) to increase the solubility of the recombinant protein; and (iii) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa, thrombin and enterokinase. Typical fusion expression vectors include pGEX (Pharmacia Biotech Inc; Smith and Johnson, 1988. *Gene* 67: 31-40), pMAL (New England Biolabs, Beverly, Mass.) and pRIT5 (Pharmacia, Piscataway, N.J.) that fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein.

Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amrann *et al.*, (1988) *Gene* 69:301-315) and pET 11d (Studier *et al.*, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 60-89).

One strategy to maximize recombinant protein expression in *E. coli* is to express the protein in a host bacteria with an impaired capacity to proteolytically cleave the recombinant protein. See, e.g., Gottesman, GENE EXPRESSION TECHNOLOGY: METHODS IN ENZYMOLOGY 185, Academic Press, San Diego, Calif. (1990) 119-128. Another strategy is to alter the nucleic acid sequence of the nucleic acid to be inserted into an expression vector so that the individual codons for each amino acid are those preferentially utilized in *E. coli* (see, e.g., Wada, *et al.*, 1992. *Nucl. Acids Res.* 20: 2111-2118). Such alteration of nucleic acid sequences of the invention can be carried out by standard DNA synthesis techniques.

In another embodiment, the NOVX expression vector is a yeast expression vector. Examples of vectors for expression in yeast *Saccharomyces cerevisiae* include pYepSec I (Baldari, *et al.*, 1987. *EMBO J.* 6: 229-234), pMFa (Kurjan and Herskowitz, 1982. *Cell* 30:

933-943), pJRY88 (Schultz *et al.*, 1987. *Gene* 54: 113-123), pYES2 (Invitrogen Corporation, San Diego, Calif.), and picZ (Invitrogen Corp, San Diego, Calif.).

Alternatively, NOVX can be expressed in insect cells using baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (*e.g.*,
5 SF9 cells) include the pAc series (Smith, *et al.*, 1983. *Mol. Cell. Biol.* 3: 2156-2165) and the pVL series (Lucklow and Summers, 1989. *Virology* 170: 31-39).

In yet another embodiment, a nucleic acid of the invention is expressed in mammalian cells using a mammalian expression vector. Examples of mammalian expression vectors include pCDM8 (Seed, 1987. *Nature* 329: 840) and pMT2PC (Kaufman, *et al.*, 1987. *EMBO J.* 6: 187-195).
10 When used in mammalian cells, the expression vector's control functions are often provided by viral regulatory elements. For example, commonly used promoters are derived from polyoma, adenovirus 2, cytomegalovirus, and simian virus 40. For other suitable expression systems for both prokaryotic and eukaryotic cells see, *e.g.*, Chapters 16 and 17 of Sambrook, *et al.*, *MOLECULAR CLONING: A LABORATORY MANUAL*. 2nd ed., Cold
15 Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1989.

In another embodiment, the recombinant mammalian expression vector is capable of directing expression of the nucleic acid preferentially in a particular cell type (*e.g.*, tissue-specific regulatory elements are used to express the nucleic acid). Tissue-specific
20 regulatory elements are known in the art. Non-limiting examples of suitable tissue-specific promoters include the albumin promoter (liver-specific; Pinkert, *et al.*, 1987. *Genes Dev.* 1: 268-277), lymphoid-specific promoters (Calame and Eaton, 1988. *Adv. Immunol.* 43: 235-275), in particular promoters of T cell receptors (Winoto and Baltimore, 1989. *EMBO J.* 8: 729-733) and immunoglobulins (Banerji, *et al.*, 1983. *Cell* 33: 729-740; Queen and
25 Baltimore, 1983. *Cell* 33: 741-748), neuron-specific promoters (*e.g.*, the neurofilament promoter; Byrne and Ruddle, 1989. *Proc. Natl. Acad. Sci. USA* 86: 5473-5477), pancreas-specific promoters (Edlund, *et al.*, 1985. *Science* 230: 912-916), and mammary gland-specific promoters (*e.g.*, milk whey promoter; U.S. Pat. No. 4,873,316 and European Application Publication No. 264,166). Developmentally-regulated promoters are also
30 encompassed, *e.g.*, the murine hox promoters (Kessel and Gruss, 1990. *Science* 249: 374-379) and the α -fetoprotein promoter (Campes and Tilghman, 1989. *Genes Dev.* 3: 537-546).

The invention further provides a recombinant expression vector comprising a DNA molecule of the invention cloned into the expression vector in an antisense orientation. That is, the DNA molecule is operatively-linked to a regulatory sequence in a manner that allows for expression (by transcription of the DNA molecule) of an RNA molecule that is antisense to NOVX mRNA. Regulatory sequences operatively linked to a nucleic acid cloned in the antisense orientation can be chosen that direct the continuous expression of the antisense RNA molecule in a variety of cell types, for instance viral promoters and/or enhancers, or regulatory sequences can be chosen that direct constitutive, tissue specific or cell type specific expression of antisense RNA. The antisense expression vector can be in the form of a recombinant plasmid, phagemid or attenuated virus in which antisense nucleic acids are produced under the control of a high efficiency regulatory region, the activity of which can be determined by the cell type into which the vector is introduced. For a discussion of the regulation of gene expression using antisense genes *see, e.g.,* Weintraub, *et al.*, "Antisense RNA as a molecular tool for genetic analysis," *Reviews-Trends in Genetics*, Vol. 1(1) 1986.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, NOVX protein can be expressed in bacterial cells such as *E. coli*, insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (*e.g.,* DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (MOLECULAR CLONING: A LABORATORY MANUAL. 2nd ed., Cold Spring

Harbor Laboratory, Cold Spring Harbor Laboratory Press. Cold Spring Harbor, N.Y., 1989), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Various selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding NOVX or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) NOVX protein. Accordingly, the invention further provides methods for producing NOVX protein using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding NOVX protein has been introduced) in a suitable medium such that NOVX protein is produced. In another embodiment, the method further comprises isolating NOVX protein from the medium or the host cell.

20 **Transgenic NOVX Animals**

The host cells of the invention can also be used to produce non-human transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which NOVX protein-coding sequences have been introduced. Such host cells can then be used to create non-human transgenic animals in which exogenous NOVX sequences have been introduced into their genome or homologous recombinant animals in which endogenous NOVX sequences have been altered. Such animals are useful for studying the function and/or activity of NOVX protein and for identifying and/or evaluating modulators of NOVX protein activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, amphibians, *etc.* A transgene is exogenous DNA that is integrated into the genome of a cell

from which a transgenic animal develops and that remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, a "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous
5 NOVX gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

A transgenic animal of the invention can be created by introducing NOVX-encoding nucleic acid into the male pronuclei of a fertilized oocyte (*e.g.*, by microinjection, retroviral
10 infection) and allowing the oocyte to develop in a pseudopregnant female foster animal. The human NOVX cDNA sequences, *i.e.*, any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, can be introduced as a transgene into the genome of a non-human animal. Alternatively, a non-human homologue of the human NOVX gene, such as a mouse NOVX gene, can be isolated based on hybridization to the human NOVX cDNA (described further
15 *supra*) and used as a transgene. Intronic sequences and polyadenylation signals can also be included in the transgene to increase the efficiency of expression of the transgene. A tissue-specific regulatory sequence(s) can be operably-linked to the NOVX transgene to direct expression of NOVX protein to particular cells. Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have
20 become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866; 4,870,009; and 4,873,191; and Hogan, 1986. In: MANIPULATING THE MOUSE EMBRYO, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the NOVX transgene in its genome and/or expression of NOVX mRNA
25 in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene-encoding NOVX protein can further be bred to other transgenic animals carrying other transgenes.

To create a homologous recombinant animal, a vector is prepared which contains at
30 least a portion of a NOVX gene into which a deletion, addition or substitution has been introduced to thereby alter, *e.g.*, functionally disrupt, the NOVX gene. The NOVX gene can be a human gene (*e.g.*, the cDNA of any one of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172), but more preferably, is a non-human homologue of a human NOVX

gene. For example, a mouse homologue of human NOVX gene of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, can be used to construct a homologous recombination vector suitable for altering an endogenous NOVX gene in the mouse genome. In one embodiment, the vector is designed such that, upon homologous recombination, the endogenous NOVX gene is functionally disrupted (*i.e.*, no longer encodes a functional protein; also referred to as a "knock out" vector).

Alternatively, the vector can be designed such that, upon homologous recombination, the endogenous NOVX gene is mutated or otherwise altered but still encodes functional protein (*e.g.*, the upstream regulatory region can be altered to thereby alter the expression of the endogenous NOVX protein). In the homologous recombination vector, the altered portion of the NOVX gene is flanked at its 5'- and 3'-termini by additional nucleic acid of the NOVX gene to allow for homologous recombination to occur between the exogenous NOVX gene carried by the vector and an endogenous NOVX gene in an embryonic stem cell. The additional flanking NOVX nucleic acid is of sufficient length for successful homologous recombination with the endogenous gene. Typically, several kilobases of flanking DNA (both at the 5'- and 3'-termini) are included in the vector. *See, e.g.*, Thomas, *et al.*, 1987. *Cell* 51: 503 for a description of homologous recombination vectors. The vector is then introduced into an embryonic stem cell line (*e.g.*, by electroporation) and cells in which the introduced NOVX gene has homologously-recombined with the endogenous NOVX gene are selected. *See, e.g.*, Li, *et al.*, 1992. *Cell* 69: 915.

The selected cells are then injected into a blastocyst of an animal (*e.g.*, a mouse) to form aggregation chimeras. *See, e.g.*, Bradley, 1987. In: TERATOCARCINOMAS AND EMBRYONIC STEM CELLS: A PRACTICAL APPROACH, Robertson, ed. IRL, Oxford, pp. 113-152. A chimeric embryo can then be implanted into a suitable pseudopregnant female foster animal and the embryo brought to term. Progeny harboring the homologously-recombined DNA in their germ cells can be used to breed animals in which all cells of the animal contain the homologously-recombined DNA by germline transmission of the transgene. Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, 1991. *Curr. Opin. Biotechnol.* 2: 823-829; PCT International Publication Nos.: WO 90/11354; WO 91/01140; WO 92/0968; and WO 93/04169.

In another embodiment, transgenic non-humans animals can be produced that contain selected systems that allow for regulated expression of the transgene. One example of such a

system is the cre/loxP recombinase system of bacteriophage P1. For a description of the cre/loxP recombinase system, *See, e.g., Lakso, et al., 1992. Proc. Natl. Acad. Sci. USA 89: 6232-6236.* Another example of a recombinase system is the FLP recombinase system of *Saccharomyces cerevisiae*. *See, O'Gorman, et al., 1991. Science 251:1351-1355.* If a
5 cre/loxP recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the Cre recombinase and a selected protein are required. Such animals can be provided through the construction of "double" transgenic animals, *e.g.,* by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

10 Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, *et al., 1997. Nature 385: 810-813.* In brief, a cell (*e.g.,* a somatic cell) from the transgenic animal can be isolated and induced to exit the growth cycle and enter G₀ phase. The quiescent cell can then be fused, *e.g.,* through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which
15 the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyte and then transferred to pseudopregnant female foster animal. The offspring borne of this female foster animal will be a clone of the animal from which the cell (*e.g.,* the somatic cell) is isolated.

Pharmaceutical Compositions

20 The NOVX nucleic acid molecules, NOVX proteins, and anti-NOVX antibodies (also referred to herein as "active compounds") of the invention, and derivatives, fragments, analogs and homologs thereof, can be incorporated into pharmaceutical compositions suitable for administration. Such compositions typically comprise the nucleic acid molecule, protein, or antibody and a pharmaceutically acceptable carrier. As used herein, "pharmaceutically
25 acceptable carrier" is intended to include any and all solvents, dispersion media, coatings, antibacterial and antifungal agents, isotonic and absorption delaying agents, and the like, compatible with pharmaceutical administration. Suitable carriers are described in the most recent edition of Remington's Pharmaceutical Sciences, a standard reference text in the field, which is incorporated herein by reference. Preferred examples of such carriers or diluents
30 include, but are not limited to, water, saline, finger's solutions, dextrose solution, and 5% human serum albumin. Liposomes and non-aqueous vehicles such as fixed oils may also be used. The use of such media and agents for pharmaceutically active substances is well

known in the art. Except insofar as any conventional media or agent is incompatible with the active compound, use thereof in the compositions is contemplated. Supplementary active compounds can also be incorporated into the compositions.

A pharmaceutical composition of the invention is formulated to be compatible with its intended route of administration. Examples of routes of administration include parenteral, *e.g.*, intravenous, intradermal, subcutaneous, oral (*e.g.*, inhalation), transdermal (*i.e.*, topical), transmucosal, and rectal administration. Solutions or suspensions used for parenteral, intradermal, or subcutaneous application can include the following components: a sterile diluent such as water for injection, saline solution, fixed oils, polyethylene glycols, glycerine, propylene glycol or other synthetic solvents; antibacterial agents such as benzyl alcohol or methyl parabens; antioxidants such as ascorbic acid or sodium bisulfite; chelating agents such as ethylenediaminetetraacetic acid (EDTA); buffers such as acetates, citrates or phosphates, and agents for the adjustment of tonicity such as sodium chloride or dextrose. The pH can be adjusted with acids or bases, such as hydrochloric acid or sodium hydroxide. The parenteral preparation can be enclosed in ampoules, disposable syringes or multiple dose vials made of glass or plastic.

Pharmaceutical compositions suitable for injectable use include sterile aqueous solutions (where water soluble) or dispersions and sterile powders for the extemporaneous preparation of sterile injectable solutions or dispersion. For intravenous administration, suitable carriers include physiological saline, bacteriostatic water, Cremophor EL™ (BASF, Parsippany, N.J.) or phosphate buffered saline (PBS). In all cases, the composition must be sterile and should be fluid to the extent that easy syringeability exists. It must be stable under the conditions of manufacture and storage and must be preserved against the contaminating action of microorganisms such as bacteria and fungi. The carrier can be a solvent or dispersion medium containing, for example, water, ethanol, polyol (for example, glycerol, propylene glycol, and liquid polyethylene glycol, and the like), and suitable mixtures thereof. The proper fluidity can be maintained, for example, by the use of a coating such as lecithin, by the maintenance of the required particle size in the case of dispersion and by the use of surfactants. Prevention of the action of microorganisms can be achieved by various antibacterial and antifungal agents, for example, parabens, chlorobutanol, phenol, ascorbic acid, thimerosal, and the like. In many cases, it will be preferable to include isotonic agents, for example, sugars, polyalcohols such as manitol, sorbitol, sodium chloride in the composition. Prolonged absorption of the injectable compositions can be brought about by

including in the composition an agent which delays absorption, for example, aluminum monostearate and gelatin.

Sterile injectable solutions can be prepared by incorporating the active compound (e.g., a NOVX protein or anti-NOVX antibody) in the required amount in an appropriate solvent with one or a combination of ingredients enumerated above, as required, followed by filtered sterilization. Generally, dispersions are prepared by incorporating the active compound into a sterile vehicle that contains a basic dispersion medium and the required other ingredients from those enumerated above. In the case of sterile powders for the preparation of sterile injectable solutions, methods of preparation are vacuum drying and freeze-drying that yields a powder of the active ingredient plus any additional desired ingredient from a previously sterile-filtered solution thereof.

Oral compositions generally include an inert diluent or an edible carrier. They can be enclosed in gelatin capsules or compressed into tablets. For the purpose of oral therapeutic administration, the active compound can be incorporated with excipients and used in the form of tablets, troches, or capsules. Oral compositions can also be prepared using a fluid carrier for use as a mouthwash, wherein the compound in the fluid carrier is applied orally and swished and expectorated or swallowed. Pharmaceutically compatible binding agents, and/or adjuvant materials can be included as part of the composition. The tablets, pills, capsules, troches and the like can contain any of the following ingredients, or compounds of a similar nature: a binder such as microcrystalline cellulose, gum tragacanth or gelatin; an excipient such as starch or lactose, a disintegrating agent such as alginic acid, Primogel, or corn starch; a lubricant such as magnesium stearate or Sterotes; a glidant such as colloidal silicon dioxide; a sweetening agent such as sucrose or saccharin; or a flavoring agent such as peppermint, methyl salicylate, or orange flavoring.

For administration by inhalation, the compounds are delivered in the form of an aerosol spray from pressured container or dispenser which contains a suitable propellant, e.g., a gas such as carbon dioxide, or a nebulizer.

Systemic administration can also be by transmucosal or transdermal means. For transmucosal or transdermal administration, penetrants appropriate to the barrier to be permeated are used in the formulation. Such penetrants are generally known in the art, and include, for example, for transmucosal administration, detergents, bile salts, and fusidic acid derivatives. Transmucosal administration can be accomplished through the use of nasal

sprays or suppositories. For transdermal administration, the active compounds are formulated into ointments, salves, gels, or creams as generally known in the art.

The compounds can also be prepared in the form of suppositories (*e.g.*, with conventional suppository bases such as cocoa butter and other glycerides) or retention
5 enemas for rectal delivery.

In one embodiment, the active compounds are prepared with carriers that will protect the compound against rapid elimination from the body, such as a controlled release formulation, including implants and microencapsulated delivery systems. Biodegradable, biocompatible polymers can be used, such as ethylene vinyl acetate, polyanhydrides,
10 polyglycolic acid, collagen, polyorthoesters, and polylactic acid. Methods for preparation of such formulations will be apparent to those skilled in the art. The materials can also be obtained commercially from Alza Corporation and Nova Pharmaceuticals, Inc. Liposomal suspensions (including liposomes targeted to infected cells with monoclonal antibodies to viral antigens) can also be used as pharmaceutically acceptable carriers. These can be
15 prepared according to methods known to those skilled in the art, for example, as described in U.S. Patent No. 4,522,811.

It is especially advantageous to formulate oral or parenteral compositions in dosage unit form for ease of administration and uniformity of dosage. Dosage unit form as used herein refers to physically discrete units suited as unitary dosages for the subject to be
20 treated; each unit containing a predetermined quantity of active compound calculated to produce the desired therapeutic effect in association with the required pharmaceutical carrier. The specification for the dosage unit forms of the invention are dictated by and directly dependent on the unique characteristics of the active compound and the particular therapeutic effect to be achieved, and the limitations inherent in the art of compounding such an active
25 compound for the treatment of individuals.

The nucleic acid molecules of the invention can be inserted into vectors and used as gene therapy vectors. Gene therapy vectors can be delivered to a subject by, for example, intravenous injection, local administration (*see, e.g.*, U.S. Patent No. 5,328,470) or by stereotactic injection (*see, e.g.*, Chen, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91:
30 3054-3057). The pharmaceutical preparation of the gene therapy vector can include the gene therapy vector in an acceptable diluent, or can comprise a slow release matrix in which the gene delivery vehicle is imbedded. Alternatively, where the complete gene delivery vector

can be produced intact from recombinant cells, *e.g.*, retroviral vectors, the pharmaceutical preparation can include one or more cells that produce the gene delivery system.

The pharmaceutical compositions can be included in a container, pack, or dispenser together with instructions for administration.

5 Screening and Detection Methods

The isolated nucleic acid molecules of the invention can be used to express NOVX protein (*e.g.*, via a recombinant expression vector in a host cell in gene therapy applications), to detect NOVX mRNA (*e.g.*, in a biological sample) or a genetic lesion in a NOVX gene, and to modulate NOVX activity, as described further, below. In addition, the NOVX proteins
10 can be used to screen drugs or compounds that modulate the NOVX protein activity or expression as well as to treat disorders characterized by insufficient or excessive production of NOVX protein or production of NOVX protein forms that have decreased or aberrant activity compared to NOVX wild-type protein (*e.g.*; diabetes (regulates insulin release); obesity (binds and transport lipids); metabolic disturbances associated with obesity, the
15 metabolic syndrome X as well as anorexia and wasting disorders associated with chronic diseases and various cancers, and infectious disease (possesses anti-microbial activity) and the various dyslipidemias. In addition, the anti-NOVX antibodies of the invention can be used to detect and isolate NOVX proteins and modulate NOVX activity. In yet a further aspect, the invention can be used in methods to influence appetite, absorption of nutrients and the
20 disposition of metabolic substrates in both a positive and negative fashion.

The invention further pertains to novel agents identified by the screening assays described herein and uses thereof for treatments as described, *supra*.

Screening Assays

The invention provides a method (also referred to herein as a "screening assay") for
25 identifying modulators, *i.e.*, candidate or test compounds or agents (*e.g.*, peptides, peptidomimetics, small molecules or other drugs) that bind to NOVX proteins or have a stimulatory or inhibitory effect on, *e.g.*, NOVX protein expression or NOVX protein activity. The invention also includes compounds identified in the screening assays described herein.

In one embodiment, the invention provides assays for screening candidate or test
30 compounds which bind to or modulate the activity of the membrane-bound form of a NOVX protein or polypeptide or biologically-active portion thereof. The test compounds of the invention can be obtained using any of the numerous approaches in combinatorial library

methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the "one-bead one-compound" library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to peptide libraries, while the other four approaches are applicable to peptide, non-peptide oligomer or small molecule libraries of compounds. See, e.g., Lam, 1997. *Anticancer Drug Design* 12: 145.

A "small molecule" as used herein, is meant to refer to a composition that has a molecular weight of less than about 5 kD and most preferably less than about 4 kD. Small molecules can be, e.g., nucleic acids, peptides, polypeptides, peptidomimetics, carbohydrates, lipids or other organic or inorganic molecules. Libraries of chemical and/or biological mixtures, such as fungal, bacterial, or algal extracts, are known in the art and can be screened with any of the assays of the invention.

Examples of methods for the synthesis of molecular libraries can be found in the art, for example in: DeWitt, *et al.*, 1993. *Proc. Natl. Acad. Sci. U.S.A.* 90: 6909; Erb, *et al.*, 1994. *Proc. Natl. Acad. Sci. U.S.A.* 91: 11422; Zuckermann, *et al.*, 1994. *J. Med. Chem.* 37: 2678; Cho, *et al.*, 1993. *Science* 261: 1303; Carrell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2059; Carell, *et al.*, 1994. *Angew. Chem. Int. Ed. Engl.* 33: 2061; and Gallop, *et al.*, 1994. *J. Med. Chem.* 37: 1233.

Libraries of compounds may be presented in solution (e.g., Houghten, 1992. *Biotechniques* 13: 412-421), or on beads (Lam, 1991. *Nature* 354: 82-84), on chips (Fodor, 1993. *Nature* 364: 555-556), bacteria (Ladner, U.S. Patent No. 5,223,409), spores (Ladner, U.S. Patent 5,233,409), plasmids (Cull, *et al.*, 1992. *Proc. Natl. Acad. Sci. USA* 89: 1865-1869) or on phage (Scott and Smith, 1990. *Science* 249: 386-390; Devlin, 1990. *Science* 249: 404-406; Cwirla, *et al.*, 1990. *Proc. Natl. Acad. Sci. U.S.A.* 87: 6378-6382; Felici, 1991. *J. Mol. Biol.* 222: 301-310; Ladner, U.S. Patent No. 5,233,409.).

In one embodiment, an assay is a cell-based assay in which a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface is contacted with a test compound and the ability of the test compound to bind to a NOVX protein determined. The cell, for example, can be of mammalian origin or a yeast cell.

Determining the ability of the test compound to bind to the NOVX protein can be accomplished, for example, by coupling the test compound with a radioisotope or enzymatic label such that binding of the test compound to the NOVX protein or biologically-active portion thereof can be determined by detecting the labeled compound in a complex. For

example, test compounds can be labeled with ^{125}I , ^{35}S , ^{14}C , or ^3H , either directly or indirectly, and the radioisotope detected by direct counting of radioemission or by scintillation counting. Alternatively, test compounds can be enzymatically-labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by

5 determination of conversion of an appropriate substrate to product. In one embodiment, the assay comprises contacting a cell which expresses a membrane-bound form of NOVX protein, or a biologically-active portion thereof, on the cell surface with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX

10 protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX protein or a biologically-active portion thereof as compared to the known compound.

In another embodiment, an assay is a cell-based assay comprising contacting a cell expressing a membrane-bound form of NOVX protein, or a biologically-active portion

15 thereof, on the cell surface with a test compound and determining the ability of the test compound to modulate (*e.g.*, stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX or a biologically-active portion thereof can be accomplished, for example, by determining the ability of the NOVX protein to bind to or interact with a NOVX

20 target molecule. As used herein, a "target molecule" is a molecule with which a NOVX protein binds or interacts in nature, for example, a molecule on the surface of a cell which expresses a NOVX interacting protein, a molecule on the surface of a second cell, a molecule in the extracellular milieu, a molecule associated with the internal surface of a cell membrane or a cytoplasmic molecule. A NOVX target molecule can be a non-NOVX molecule or a

25 NOVX protein or polypeptide of the invention. In one embodiment, a NOVX target molecule is a component of a signal transduction pathway that facilitates transduction of an extracellular signal (*e.g.* a signal generated by binding of a compound to a membrane-bound NOVX molecule) through the cell membrane and into the cell. The target, for example, can be a second intercellular protein that has catalytic activity or a protein that facilitates the

30 association of downstream signaling molecules with NOVX.

Determining the ability of the NOVX protein to bind to or interact with a NOVX target molecule can be accomplished by one of the methods described above for determining direct binding. In one embodiment, determining the ability of the NOVX protein to bind to or

interact with a NOVX target molecule can be accomplished by determining the activity of the target molecule. For example, the activity of the target molecule can be determined by detecting induction of a cellular second messenger of the target (*i.e.* intracellular Ca^{2+} , diacylglycerol, IP_3 , *etc.*), detecting catalytic/enzymatic activity of the target an appropriate substrate, detecting the induction of a reporter gene (comprising a NOVX-responsive regulatory element operatively linked to a nucleic acid encoding a detectable marker, *e.g.*, luciferase), or detecting a cellular response, for example, cell survival, cellular differentiation, or cell proliferation.

In yet another embodiment, an assay of the invention is a cell-free assay comprising contacting a NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to bind to the NOVX protein or biologically-active portion thereof. Binding of the test compound to the NOVX protein can be determined either directly or indirectly as described above. In one such embodiment, the assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX to form an assay mixture, contacting the assay mixture with a test compound, and determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the test compound to preferentially bind to NOVX or biologically-active portion thereof as compared to the known compound.

In still another embodiment, an assay is a cell-free assay comprising contacting NOVX protein or biologically-active portion thereof with a test compound and determining the ability of the test compound to modulate (*e.g.* stimulate or inhibit) the activity of the NOVX protein or biologically-active portion thereof. Determining the ability of the test compound to modulate the activity of NOVX can be accomplished, for example, by determining the ability of the NOVX protein to bind to a NOVX target molecule by one of the methods described above for determining direct binding. In an alternative embodiment, determining the ability of the test compound to modulate the activity of NOVX protein can be accomplished by determining the ability of the NOVX protein further modulate a NOVX target molecule. For example, the catalytic/enzymatic activity of the target molecule on an appropriate substrate can be determined as described, *supra*.

In yet another embodiment, the cell-free assay comprises contacting the NOVX protein or biologically-active portion thereof with a known compound which binds NOVX protein to form an assay mixture, contacting the assay mixture with a test compound, and

determining the ability of the test compound to interact with a NOVX protein, wherein determining the ability of the test compound to interact with a NOVX protein comprises determining the ability of the NOVX protein to preferentially bind to or modulate the activity of a NOVX target molecule.

5 The cell-free assays of the invention are amenable to use of both the soluble form or the membrane-bound form of NOVX protein. In the case of cell-free assays comprising the membrane-bound form of NOVX protein, it may be desirable to utilize a solubilizing agent such that the membrane-bound form of NOVX protein is maintained in solution. Examples of such solubilizing agents include non-ionic detergents such as n-octylglucoside,
10 n-dodecylglucoside, n-dodecylmaltoside, octanoyl-N-methylglucamide, decanoyl-N-methylglucamide, Triton® X-100, Triton® X-114, Thesit®, Isotridecypoly(ethylene glycol ether)_n, N-dodecyl-N,N-dimethyl-3-ammonio-1-propane sulfonate, 3-(3-cholamidopropyl) dimethylamminiol-1-propane sulfonate (CHAPS), or 3-(3-cholamidopropyl)dimethylamminiol-2-hydroxy-1-propane sulfonate (CHAPSO).⁽²⁾

15 In more than one embodiment of the above assay methods of the invention, it may be desirable to immobilize either NOVX protein or its target molecule to facilitate separation of complexed from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay. Binding of a test compound to NOVX protein, or interaction of NOVX protein with a target molecule in the presence and absence of a
20 candidate compound, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtiter plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein can be provided that adds a domain that allows one or both of the proteins to be bound to a matrix. For example, GST-NOVX fusion proteins or GST-target fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma
25 Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, that are then combined with the test compound or the test compound and either the non-adsorbed target protein or NOVX protein, and the mixture is incubated under conditions conducive to complex formation (*e.g.*, at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components, the matrix
30 immobilized in the case of beads, complex determined either directly or indirectly, for example, as described, *supra*. Alternatively, the complexes can be dissociated from the matrix, and the level of NOVX protein binding or activity determined using standard techniques.

Other techniques for immobilizing proteins on matrices can also be used in the screening assays of the invention. For example, either the NOVX protein or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated NOVX protein or target molecules can be prepared from biotin-NHS

5 (N-hydroxy-succinimide) using techniques well-known within the art (*e.g.*, biotinylation kit, Pierce Chemicals, Rockford, Ill.), and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemical). Alternatively, antibodies reactive with NOVX protein or target molecules, but which do not interfere with binding of the NOVX protein to its target molecule, can be derivatized to the wells of the plate, and unbound target or NOVX protein
10 trapped in the wells by antibody conjugation. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the NOVX protein or target molecule, as well as enzyme-linked assays that rely on detecting an enzymatic activity associated with the NOVX protein or target molecule.

15 In another embodiment, modulators of NOVX protein expression are identified in a method wherein a cell is contacted with a candidate compound and the expression of NOVX mRNA or protein in the cell is determined. The level of expression of NOVX mRNA or protein in the presence of the candidate compound is compared to the level of expression of NOVX mRNA or protein in the absence of the candidate compound. The candidate
20 compound can then be identified as a modulator of NOVX mRNA or protein expression based upon this comparison. For example, when expression of NOVX mRNA or protein is greater (*i.e.*, statistically significantly greater) in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of NOVX mRNA or protein expression. Alternatively, when expression of NOVX mRNA or protein is less
25 (statistically significantly less) in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of NOVX mRNA or protein expression. The level of NOVX mRNA or protein expression in the cells can be determined by methods described herein for detecting NOVX mRNA or protein.

In yet another aspect of the invention, the NOVX proteins can be used as "bait
30 proteins" in a two-hybrid assay or three hybrid assay (*see, e.g.*, U.S. Patent No. 5,283,317; Zervos, *et al.*, 1993. *Cell* 72: 223-232; Madura, *et al.*, 1993. *J. Biol. Chem.* 268: 12046-12054; Bartel, *et al.*, 1993. *Biotechniques* 14: 920-924; Iwabuchi, *et al.*, 1993. *Oncogene* 8: 1693-1696; and Brent WO 94/10300), to identify other proteins that bind to or

interact with NOVX ("NOVX-binding proteins" or "NOVX-bp") and modulate NOVX activity. Such NOVX-binding proteins are also involved in the propagation of signals by the NOVX proteins as, for example, upstream or downstream elements of the NOVX pathway.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for NOVX is fused to a gene encoding the DNA binding domain of a known transcription factor (*e.g.*, GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo*, forming a NOVX-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (*e.g.*, LacZ) that is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene that encodes the protein which interacts with NOVX.

The invention further pertains to novel agents identified by the aforementioned screening assays and uses thereof for treatments as described herein.

Detection Assays

Portions or fragments of the cDNA sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. By way of example, and not of limitation, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Some of these applications are described in the subsections, below.

Chromosome Mapping

Once the sequence (or a portion of the sequence) of a gene has been isolated, this sequence can be used to map the location of the gene on a chromosome. This process is called chromosome mapping. Accordingly, portions or fragments of the NOVX sequences of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or fragments or derivatives thereof, can be used to map the location of the NOVX genes, respectively, on a chromosome.

The mapping of the NOVX sequences to chromosomes is an important first step in correlating these sequences with genes associated with disease.

Briefly, NOVX genes can be mapped to chromosomes by preparing PCR primers (preferably 15-25 bp in length) from the NOVX sequences. Computer analysis of the NOVX sequences can be used to rapidly select primers that do not span more than one exon in the genomic DNA, thus complicating the amplification process. These primers can then be used for PCR screening of somatic cell hybrids containing individual human chromosomes. Only those hybrids containing the human gene corresponding to the NOVX sequences will yield an amplified fragment.

Somatic cell hybrids are prepared by fusing somatic cells from different mammals (e.g., human and mouse cells). As hybrids of human and mouse cells grow and divide, they gradually lose human chromosomes in random order, but retain the mouse chromosomes. By using media in which mouse cells cannot grow, because they lack a particular enzyme, but in which human cells can, the one human chromosome that contains the gene encoding the needed enzyme will be retained. By using various media, panels of hybrid cell lines can be established. Each cell line in a panel contains either a single human chromosome or a small number of human chromosomes, and a full set of mouse chromosomes, allowing easy mapping of individual genes to specific human chromosomes. See, e.g., D'Eustachio, *et al.*, 1983. *Science* 220: 919-924. Somatic cell hybrids containing only fragments of human chromosomes can also be produced by using human chromosomes with translocations and deletions.

PCR mapping of somatic cell hybrids is a rapid procedure for assigning a particular sequence to a particular chromosome. Three or more sequences can be assigned per day using a single thermal cycler. Using the NOVX sequences to design oligonucleotide primers, sub-localization can be achieved with panels of fragments from specific chromosomes.

Fluorescence *in situ* hybridization (FISH) of a DNA sequence to a metaphase chromosomal spread can further be used to provide a precise chromosomal location in one step. Chromosome spreads can be made using cells whose division has been blocked in metaphase by a chemical like colcemid that disrupts the mitotic spindle. The chromosomes can be treated briefly with trypsin, and then stained with Giemsa. A pattern of light and dark bands develops on each chromosome, so that the chromosomes can be identified individually. The FISH technique can be used with a DNA sequence as short as 500 or 600 bases. However, clones larger than 1,000 bases have a higher likelihood of binding to a unique

chromosomal location with sufficient signal intensity for simple detection. Preferably 1,000 bases, and more preferably 2,000 bases, will suffice to get good results at a reasonable amount of time. For a review of this technique, *see*, Verma, *et al.*, HUMAN CHROMOSOMES: A MANUAL OF BASIC TECHNIQUES (Pergamon Press, New York 1988).

5 Reagents for chromosome mapping can be used individually to mark a single chromosome or a single site on that chromosome, or panels of reagents can be used for marking multiple sites and/or multiple chromosomes. Reagents corresponding to noncoding regions of the genes actually are preferred for mapping purposes. Coding sequences are more likely to be conserved within gene families, thus increasing the chance of cross hybridizations
10 during chromosomal mapping.

 Once a sequence has been mapped to a precise chromosomal location, the physical position of the sequence on the chromosome can be correlated with genetic map data. Such data are found, *e.g.*, in McKusick, MENDELIAN INHERITANCE IN MAN, available on-line through Johns Hopkins University Welch Medical Library). The relationship between genes
15 and disease, mapped to the same chromosomal region, can then be identified through linkage analysis (co-inheritance of physically adjacent genes), described in, *e.g.*, Egeland, *et al.*, 1987. *Nature*, 325: 783-787.

 Moreover, differences in the DNA sequences between individuals affected and unaffected with a disease associated with the NOVX gene, can be determined. If a mutation
20 is observed in some or all of the affected individuals but not in any unaffected individuals, then the mutation is likely to be the causative agent of the particular disease. Comparison of affected and unaffected individuals generally involves first looking for structural alterations in the chromosomes, such as deletions or translocations that are visible from chromosome spreads or detectable using PCR based on that DNA sequence. Ultimately, complete
25 sequencing of genes from several individuals can be performed to confirm the presence of a mutation and to distinguish mutations from polymorphisms.

Tissue Typing

 The NOVX sequences of the invention can also be used to identify individuals from minute biological samples. In this technique, an individual's genomic DNA is digested with
30 one or more restriction enzymes, and probed on a Southern blot to yield unique bands for identification. The sequences of the invention are useful as additional DNA markers for

RFLP ("restriction fragment length polymorphisms," described in U.S. Patent No. 5,272,057).

Furthermore, the sequences of the invention can be used to provide an alternative technique that determines the actual base-by-base DNA sequence of selected portions of an individual's genome. Thus, the NOVX sequences described herein can be used to prepare two PCR primers from the 5'- and 3'-termini of the sequences. These primers can then be used to amplify an individual's DNA and subsequently sequence it.

Panels of corresponding DNA sequences from individuals, prepared in this manner, can provide unique individual identifications, as each individual will have a unique set of such DNA sequences due to allelic differences. The sequences of the invention can be used to obtain such identification sequences from individuals and from tissue. The NOVX sequences of the invention uniquely represent portions of the human genome. Allelic variation occurs to some degree in the coding regions of these sequences, and to a greater degree in the noncoding regions. It is estimated that allelic variation between individual humans occurs with a frequency of about once per each 500 bases. Much of the allelic variation is due to single nucleotide polymorphisms (SNPs), which include restriction fragment length polymorphisms (RFLPs).

Each of the sequences described herein can, to some degree, be used as a standard against which DNA from an individual can be compared for identification purposes. Because greater numbers of polymorphisms occur in the noncoding regions, fewer sequences are necessary to differentiate individuals. The noncoding sequences can comfortably provide positive individual identification with a panel of perhaps 10 to 1,000 primers that each yield a noncoding amplified sequence of 100 bases. If coding sequences, such as those of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, are used, a more appropriate number of primers for positive individual identification would be 500-2,000.

Predictive Medicine

The invention also pertains to the field of predictive medicine in which diagnostic assays, prognostic assays, pharmacogenomics, and monitoring clinical trials are used for prognostic (predictive) purposes to thereby treat an individual prophylactically. Accordingly, one aspect of the invention relates to diagnostic assays for determining NOVX protein and/or nucleic acid expression as well as NOVX activity, in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a

disease or disorder, or is at risk of developing a disorder, associated with aberrant NOVX expression or activity. The disorders include metabolic disorders, diabetes, obesity, infectious disease, anorexia, cancer-associated cachexia, cancer, neurodegenerative disorders, Alzheimer's Disease, Parkinson's Disorder, immune disorders, and hematopoietic disorders, and the various dyslipidemias, metabolic disturbances associated with obesity, the metabolic syndrome X and wasting disorders associated with chronic diseases and various cancers. The invention also provides for prognostic (or predictive) assays for determining whether an individual is at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. For example, mutations in a NOVX gene can be assayed in a biological sample. Such assays can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of a disorder characterized by or associated with NOVX protein, nucleic acid expression, or biological activity.

Another aspect of the invention provides methods for determining NOVX protein, nucleic acid expression or activity in an individual to thereby select appropriate therapeutic or prophylactic agents for that individual (referred to herein as "pharmacogenomics"). Pharmacogenomics allows for the selection of agents (*e.g.*, drugs) for therapeutic or prophylactic treatment of an individual based on the genotype of the individual (*e.g.*, the genotype of the individual examined to determine the ability of the individual to respond to a particular agent.)

Yet another aspect of the invention pertains to monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or activity of NOVX in clinical trials.

These and other agents are described in further detail in the following sections.

Diagnostic Assays

An exemplary method for detecting the presence or absence of NOVX in a biological sample involves obtaining a biological sample from a test subject and contacting the biological sample with a compound or an agent capable of detecting NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) that encodes NOVX protein such that the presence of NOVX is detected in the biological sample. An agent for detecting NOVX mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to NOVX mRNA or genomic DNA. The nucleic acid probe can be, for example, a full-length NOVX nucleic acid, such as the nucleic acid of SEQ ID NO:2*n*-1, wherein *n* is an integer between 1 and 172, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500

nucleotides in length and sufficient to specifically hybridize under stringent conditions to NOVX mRNA or genomic DNA. Other suitable probes for use in the diagnostic assays of the invention are described herein.

An agent for detecting NOVX protein is an antibody capable of binding to NOVX protein, preferably an antibody with a detectable label. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or F(ab')₂) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*, physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently-labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently-labeled streptavidin. The term "biological sample" is intended to include tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject. That is, the detection method of the invention can be used to detect NOVX mRNA, protein, or genomic DNA in a biological sample *in vitro* as well as *in vivo*. For example, *in vitro* techniques for detection of NOVX mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detection of NOVX protein include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations, and immunofluorescence. *In vitro* techniques for detection of NOVX genomic DNA include Southern hybridizations. Furthermore, *in vivo* techniques for detection of NOVX protein include introducing into a subject a labeled anti-NOVX antibody. For example, the antibody can be labeled with a radioactive marker whose presence and location in a subject can be detected by standard imaging techniques.

In one embodiment, the biological sample contains protein molecules from the test subject. Alternatively, the biological sample can contain mRNA molecules from the test subject or genomic DNA molecules from the test subject. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject.

In another embodiment, the methods further involve obtaining a control biological sample from a control subject, contacting the control sample with a compound or agent capable of detecting NOVX protein, mRNA, or genomic DNA, such that the presence of NOVX protein, mRNA or genomic DNA is detected in the biological sample, and comparing

the presence of NOVX protein, mRNA or genomic DNA in the control sample with the presence of NOVX protein, mRNA or genomic DNA in the test sample.

The invention also encompasses kits for detecting the presence of NOVX in a biological sample. For example, the kit can comprise: a labeled compound or agent capable of detecting NOVX protein or mRNA in a biological sample; means for determining the amount of NOVX in the sample; and means for comparing the amount of NOVX in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect NOVX protein or nucleic acid.

10 **Prognostic Assays**

The diagnostic methods described herein can furthermore be utilized to identify subjects having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. For example, the assays described herein, such as the preceding diagnostic assays or the following assays, can be utilized to identify a subject having or at risk of developing a disorder associated with NOVX protein, nucleic acid expression or activity. Alternatively, the prognostic assays can be utilized to identify a subject having or at risk for developing a disease or disorder. Thus, the invention provides a method for identifying a disease or disorder associated with aberrant NOVX expression or activity in which a test sample is obtained from a subject and NOVX protein or nucleic acid (*e.g.*, mRNA, genomic DNA) is detected, wherein the presence of NOVX protein or nucleic acid is diagnostic for a subject having or at risk of developing a disease or disorder associated with aberrant NOVX expression or activity. As used herein, a "test sample" refers to a biological sample obtained from a subject of interest. For example, a test sample can be a biological fluid (*e.g.*, serum), cell sample, or tissue.

Furthermore, the prognostic assays described herein can be used to determine whether a subject can be administered an agent (*e.g.*, an agonist, antagonist, peptidomimetic, protein, peptide, nucleic acid, small molecule, or other drug candidate) to treat a disease or disorder associated with aberrant NOVX expression or activity. For example, such methods can be used to determine whether a subject can be effectively treated with an agent for a disorder. Thus, the invention provides methods for determining whether a subject can be effectively treated with an agent for a disorder associated with aberrant NOVX expression or activity in which a test sample is obtained and NOVX protein or nucleic acid is detected (*e.g.*, wherein

the presence of NOVX protein or nucleic acid is diagnostic for a subject that can be administered the agent to treat a disorder associated with aberrant NOVX expression or activity).

The methods of the invention can also be used to detect genetic lesions in a NOVX gene, thereby determining if a subject with the lesioned gene is at risk for a disorder characterized by aberrant cell proliferation and/or differentiation. In various embodiments, the methods include detecting, in a sample of cells from the subject, the presence or absence of a genetic lesion characterized by at least one of an alteration affecting the integrity of a gene encoding a NOVX-protein, or the misexpression of the NOVX gene. For example, such genetic lesions can be detected by ascertaining the existence of at least one of: (i) a deletion of one or more nucleotides from a NOVX gene; (ii) an addition of one or more nucleotides to a NOVX gene; (iii) a substitution of one or more nucleotides of a NOVX gene, (iv) a chromosomal rearrangement of a NOVX gene; (v) an alteration in the level of a messenger RNA transcript of a NOVX gene, (vi) aberrant modification of a NOVX gene, such as of the methylation pattern of the genomic DNA, (vii) the presence of a non-wild-type splicing pattern of a messenger RNA transcript of a NOVX gene, (viii) a non-wild-type level of a NOVX protein, (ix) allelic loss of a NOVX gene, and (x) inappropriate post-translational modification of a NOVX protein. As described herein, there are a large number of assay techniques known in the art which can be used for detecting lesions in a NOVX gene. A preferred biological sample is a peripheral blood leukocyte sample isolated by conventional means from a subject. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

In certain embodiments, detection of the lesion involves the use of a probe/primer in a polymerase chain reaction (PCR) (*see, e.g.*, U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (*see, e.g.*, Landegran, *et al.*, 1988. *Science* 241: 1077-1080; and Nakazawa, *et al.*, 1994. *Proc. Natl. Acad. Sci. USA* 91: 360-364), the latter of which can be particularly useful for detecting point mutations in the NOVX-gene (*see*, Abravaya, *et al.*, 1995. *Nucl. Acids Res.* 23: 675-682). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (*e.g.*, genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers that specifically hybridize to a NOVX gene under conditions such that hybridization and amplification of the NOVX gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the

size of the amplification product and comparing the length to a control sample. It is anticipated that PCR and/or LCR may be desirable to use as a preliminary amplification step in conjunction with any of the techniques used for detecting mutations described herein.

Alternative amplification methods include: self sustained sequence replication (*see*,
5 Guatelli, *et al.*, 1990. *Proc. Natl. Acad. Sci. USA* 87: 1874-1878), transcriptional
amplification system (*see*, Kwok, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 1173-1177);
Q β Replicase (*see*, Lizardi, *et al.*, 1988. *BioTechnology* 6: 1197), or any other nucleic acid
amplification method, followed by the detection of the amplified molecules using techniques
well known to those of skill in the art. These detection schemes are especially useful for the
10 detection of nucleic acid molecules if such molecules are present in very low numbers.

In an alternative embodiment, mutations in a NOVX gene from a sample cell can be
identified by alterations in restriction enzyme cleavage patterns. For example, sample and
control DNA is isolated, amplified (optionally), digested with one or more restriction
endonucleases, and fragment length sizes are determined by gel electrophoresis and
15 compared. Differences in fragment length sizes between sample and control DNA indicates
mutations in the sample DNA. Moreover, the use of sequence specific ribozymes (*see, e.g.*,
U.S. Patent No. 5,493,531) can be used to score for the presence of specific mutations by
development or loss of a ribozyme cleavage site.

In other embodiments, genetic mutations in NOVX can be identified by hybridizing a
20 sample and control nucleic acids, *e.g.*, DNA or RNA, to high-density arrays containing
hundreds or thousands of oligonucleotides probes. *See, e.g.*, Cronin, *et al.*, 1996. *Human
Mutation* 7: 244-255; Kozal, *et al.*, 1996. *Nat. Med.* 2: 753-759. For example, genetic
mutations in NOVX can be identified in two dimensional arrays containing light-generated
DNA probes as described in Cronin, *et al., supra*. Briefly, a first hybridization array of
25 probes can be used to scan through long stretches of DNA in a sample and control to identify
base changes between the sequences by making linear arrays of sequential overlapping
probes. This step allows the identification of point mutations. This is followed by a second
hybridization array that allows the characterization of specific mutations by using smaller,
specialized probe arrays complementary to all variants or mutations detected. Each mutation
30 array is composed of parallel probe sets, one complementary to the wild-type gene and the
other complementary to the mutant gene.

In yet another embodiment, any of a variety of sequencing reactions known in the art
can be used to directly sequence the NOVX gene and detect mutations by comparing the

sequence of the sample NOVX with the corresponding wild-type (control) sequence. Examples of sequencing reactions include those based on techniques developed by Maxim and Gilbert, 1977. *Proc. Natl. Acad. Sci. USA* 74: 560 or Sanger, 1977. *Proc. Natl. Acad. Sci. USA* 74: 5463. It is also contemplated that any of a variety of automated sequencing procedures can be utilized when performing the diagnostic assays (see, e.g., Naeve, *et al.*, 1995. *Biotechniques* 19: 448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen, *et al.*, 1996. *Adv. Chromatography* 36: 127-162; and Griffin, *et al.*, 1993. *Appl. Biochem. Biotechnol.* 38: 147-159).

Other methods for detecting mutations in the NOVX gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA heteroduplexes. See, e.g., Myers, *et al.*, 1985. *Science* 230: 1242. In general, the art technique of "mismatch cleavage" starts by providing heteroduplexes of formed by hybridizing (labeled) RNA or DNA containing the wild-type NOVX sequence with potentially mutant RNA or DNA obtained from a tissue sample. The double-stranded duplexes are treated with an agent that cleaves single-stranded regions of the duplex such as which will exist due to basepair mismatches between the control and sample strands. For instance, RNA/DNA duplexes can be treated with RNase and DNA/DNA hybrids treated with S₁ nuclease to enzymatically digesting the mismatched regions. In other embodiments, either DNA/DNA or RNA/DNA duplexes can be treated with hydroxylamine or osmium tetroxide and with piperidine in order to digest mismatched regions. After digestion of the mismatched regions, the resulting material is then separated by size on denaturing polyacrylamide gels to determine the site of mutation. See, e.g., Cotton, *et al.*, 1988. *Proc. Natl. Acad. Sci. USA* 85: 4397; Saleeba, *et al.*, 1992. *Methods Enzymol.* 217: 286-295. In an embodiment, the control DNA or RNA can be labeled for detection.

In still another embodiment, the mismatch cleavage reaction employs one or more proteins that recognize mismatched base pairs in double-stranded DNA (so called "DNA mismatch repair" enzymes) in defined systems for detecting and mapping point mutations in NOVX cDNAs obtained from samples of cells. For example, the mutY enzyme of *E. coli* cleaves A at G/A mismatches and the thymidine DNA glycosylase from HeLa cells cleaves T at G/T mismatches. See, e.g., Hsu, *et al.*, 1994. *Carcinogenesis* 15: 1657-1662. According to an exemplary embodiment, a probe based on a NOVX sequence, e.g., a wild-type NOVX sequence, is hybridized to a cDNA or other DNA product from a test cell(s). The duplex is

treated with a DNA mismatch repair enzyme, and the cleavage products, if any, can be detected from electrophoresis protocols or the like. *See, e.g.*, U.S. Patent No. 5,459,039.

In other embodiments, alterations in electrophoretic mobility will be used to identify mutations in NOVX genes. For example, single strand conformation polymorphism (SSCP) may be used to detect differences in electrophoretic mobility between mutant and wild type nucleic acids. *See, e.g.*, Orita, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA*: 86: 2766; Cotton, 1993. *Mutat. Res.* 285: 125-144; Hayashi, 1992. *Genet. Anal. Tech. Appl.* 9: 73-79. Single-stranded DNA fragments of sample and control NOVX nucleic acids will be denatured and allowed to renature. The secondary structure of single-stranded nucleic acids varies according to sequence, the resulting alteration in electrophoretic mobility enables the detection of even a single base change. The DNA fragments may be labeled or detected with labeled probes. The sensitivity of the assay may be enhanced by using RNA (rather than DNA), in which the secondary structure is more sensitive to a change in sequence. In one embodiment, the subject method utilizes heteroduplex analysis to separate double stranded heteroduplex molecules on the basis of changes in electrophoretic mobility. *See, e.g.*, Keen, *et al.*, 1991. *Trends Genet.* 7: 5.

In yet another embodiment, the movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (DGGE). *See, e.g.*, Myers, *et al.*, 1985. *Nature* 313: 495. When DGGE is used as the method of analysis, DNA will be modified to insure that it does not completely denature, for example by adding a GC clamp of approximately 40 bp of high-melting GC-rich DNA by PCR. In a further embodiment, a temperature gradient is used in place of a denaturing gradient to identify differences in the mobility of control and sample DNA. *See, e.g.*, Rosenbaum and Reissner, 1987. *Biophys. Chem.* 265: 12753.

Examples of other techniques for detecting point mutations include, but are not limited to, selective oligonucleotide hybridization, selective amplification, or selective primer extension. For example, oligonucleotide primers may be prepared in which the known mutation is placed centrally and then hybridized to target DNA under conditions that permit hybridization only if a perfect match is found. *See, e.g.*, Saiki, *et al.*, 1986. *Nature* 324: 163; Saiki, *et al.*, 1989. *Proc. Natl. Acad. Sci. USA* 86: 6230. Such allele specific oligonucleotides are hybridized to PCR amplified target DNA or a number of different mutations when the oligonucleotides are attached to the hybridizing membrane and hybridized with labeled target DNA.

Alternatively, allele specific amplification technology that depends on selective PCR amplification may be used in conjunction with the instant invention. Oligonucleotides used as primers for specific amplification may carry the mutation of interest in the center of the molecule (so that amplification depends on differential hybridization; *see, e.g.*, Gibbs, *et al.*, 1989. *Nucl. Acids Res.* 17: 2437-2448) or at the extreme 3'-terminus of one primer where, under appropriate conditions, mismatch can prevent, or reduce polymerase extension (*see, e.g.*, Prossner, 1993. *Tibtech.* 11: 238). In addition it may be desirable to introduce a novel restriction site in the region of the mutation to create cleavage-based detection. *See, e.g.*, Gasparini, *et al.*, 1992. *Mol. Cell Probes* 6: 1. It is anticipated that in certain embodiments amplification may also be performed using *Taq* ligase for amplification. *See, e.g.*, Barany, 1991. *Proc. Natl. Acad. Sci. USA* 88: 189. In such cases, ligation will occur only if there is a perfect match at the 3'-terminus of the 5' sequence, making it possible to detect the presence of a known mutation at a specific site by looking for the presence or absence of amplification.

The methods described herein may be performed, for example, by utilizing pre-packaged diagnostic kits comprising at least one probe nucleic acid or antibody reagent described herein, which may be conveniently used, *e.g.*, in clinical settings to diagnose patients exhibiting symptoms or family history of a disease or illness involving a NOVX gene.

Furthermore, any cell type or tissue, preferably peripheral blood leukocytes, in which NOVX is expressed may be utilized in the prognostic assays described herein. However, any biological sample containing nucleated cells may be used, including, for example, buccal mucosal cells.

Pharmacogenomics

Agents, or modulators that have a stimulatory or inhibitory effect on NOVX activity (*e.g.*, NOVX gene expression), as identified by a screening assay described herein can be administered to individuals to treat (prophylactically or therapeutically) disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

In conjunction with such treatment, the pharmacogenomics (*i.e.*, the study of the relationship between an individual's genotype and that individual's response to a foreign compound or drug) of the individual may be considered. Differences in metabolism of

therapeutics can lead to severe toxicity or therapeutic failure by altering the relation between dose and blood concentration of the pharmacologically active drug. Thus, the pharmacogenomics of the individual permits the selection of effective agents (*e.g.*, drugs) for prophylactic or therapeutic treatments based on a consideration of the individual's genotype.

- 5 Such pharmacogenomics can further be used to determine appropriate dosages and therapeutic regimens. Accordingly, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual.

- Pharmacogenomics deals with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See *e.g.*, Eichelbaum, 1996. *Clin. Exp. Pharmacol. Physiol.*, 23: 983-985; Linder, 1997. *Clin. Chem.*, 43: 254-266. In general, two types of pharmacogenetic conditions can be differentiated. Genetic conditions transmitted as a single factor altering the way drugs act on the body (altered drug action) or genetic conditions transmitted as single factors altering the way the body acts on drugs (altered drug metabolism). These pharmacogenetic conditions can occur either as rare defects or as polymorphisms. For example, glucose-6-phosphate dehydrogenase (G6PD) deficiency is a common inherited enzymopathy in which the main clinical complication is hemolysis after ingestion of oxidant drugs (anti-malarials, sulfonamides, analgesics, nitrofurans) and consumption of fava beans.

- 20 As an illustrative embodiment, the activity of drug metabolizing enzymes is a major determinant of both the intensity and duration of drug action. The discovery of genetic polymorphisms of drug metabolizing enzymes (*e.g.*, N-acetyltransferase 2 (NAT 2) and cytochrome pregnancy zone protein precursor enzymes CYP2D6 and CYP2C19) has provided an explanation as to why some patients do not obtain the expected drug effects or show exaggerated drug response and serious toxicity after taking the standard and safe dose of a drug. These polymorphisms are expressed in two phenotypes in the population, the extensive metabolizer (EM) and poor metabolizer (PM). The prevalence of PM is different among different populations. For example, the gene coding for CYP2D6 is highly polymorphic and several mutations have been identified in PM, which all lead to the absence of functional CYP2D6. Poor metabolizers of CYP2D6 and CYP2C19 quite frequently experience exaggerated drug response and side effects when they receive standard doses. If a metabolite is the active therapeutic moiety, PM show no therapeutic response, as demonstrated for the analgesic effect of codeine mediated by its CYP2D6-formed metabolite

morphine. At the other extreme are the so called ultra-rapid metabolizers who do not respond to standard doses. Recently, the molecular basis of ultra-rapid metabolism has been identified to be due to CYP2D6 gene amplification.

Thus, the activity of NOVX protein, expression of NOVX nucleic acid, or mutation
5 content of NOVX genes in an individual can be determined to thereby select appropriate agent(s) for therapeutic or prophylactic treatment of the individual. In addition, pharmacogenetic studies can be used to apply genotyping of polymorphic alleles encoding drug-metabolizing enzymes to the identification of an individual's drug responsiveness phenotype. This knowledge, when applied to dosing or drug selection, can avoid adverse
10 reactions or therapeutic failure and thus enhance therapeutic or prophylactic efficiency when treating a subject with a NOVX modulator, such as a modulator identified by one of the exemplary screening assays described herein.

Monitoring of Effects During Clinical Trials

Monitoring the influence of agents (*e.g.*, drugs, compounds) on the expression or
15 activity of NOVX (*e.g.*, the ability to modulate aberrant cell proliferation and/or differentiation) can be applied not only in basic drug screening, but also in clinical trials. For example, the effectiveness of an agent determined by a screening assay as described herein to increase NOVX gene expression, protein levels, or upregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting decreased NOVX gene expression, protein
20 levels, or downregulated NOVX activity. Alternatively, the effectiveness of an agent determined by a screening assay to decrease NOVX gene expression, protein levels, or downregulate NOVX activity, can be monitored in clinical trials of subjects exhibiting increased NOVX gene expression, protein levels, or upregulated NOVX activity. In such clinical trials, the expression or activity of NOVX and, preferably, other genes that have been
25 implicated in, for example, a cellular proliferation or immune disorder can be used as a "read out" or markers of the immune responsiveness of a particular cell.

By way of example, and not of limitation, genes, including NOVX, that are modulated in cells by treatment with an agent (*e.g.*, compound, drug or small molecule) that modulates NOVX activity (*e.g.*, identified in a screening assay as described herein) can be
30 identified. Thus, to study the effect of agents on cellular proliferation disorders, for example, in a clinical trial, cells can be isolated and RNA prepared and analyzed for the levels of expression of NOVX and other genes implicated in the disorder. The levels of gene

expression (*i.e.*, a gene expression pattern) can be quantified by Northern blot analysis or RT-PCR, as described herein, or alternatively by measuring the amount of protein produced, by one of the methods as described herein, or by measuring the levels of activity of NOVX or other genes. In this manner, the gene expression pattern can serve as a marker, indicative of the physiological response of the cells to the agent. Accordingly, this response state may be determined before, and at various points during, treatment of the individual with the agent.

In one embodiment, the invention provides a method for monitoring the effectiveness of treatment of a subject with an agent (*e.g.*, an agonist, antagonist, protein, peptide, peptidomimetic, nucleic acid, small molecule, or other drug candidate identified by the screening assays described herein) comprising the steps of (i) obtaining a pre-administration sample from a subject prior to administration of the agent; (ii) detecting the level of expression of a NOVX protein, mRNA, or genomic DNA in the preadministration sample; (iii) obtaining one or more post-administration samples from the subject; (iv) detecting the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the post-administration samples; (v) comparing the level of expression or activity of the NOVX protein, mRNA, or genomic DNA in the pre-administration sample with the NOVX protein, mRNA, or genomic DNA in the post administration sample or samples; and (vi) altering the administration of the agent to the subject accordingly. For example, increased administration of the agent may be desirable to increase the expression or activity of NOVX to higher levels than detected, *i.e.*, to increase the effectiveness of the agent. Alternatively, decreased administration of the agent may be desirable to decrease expression or activity of NOVX to lower levels than detected, *i.e.*, to decrease the effectiveness of the agent.

Methods of Treatment

The invention provides for both prophylactic and therapeutic methods of treating a subject at risk of (or susceptible to) a disorder or having a disorder associated with aberrant NOVX expression or activity. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

These methods of treatment will be discussed more fully, below.

Diseases and Disorders

Diseases and disorders that are characterized by increased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that antagonize (*i.e.*, reduce or inhibit) activity. Therapeutics that antagonize activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to: (i) an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; (ii) antibodies to an aforementioned peptide; (iii) nucleic acids encoding an aforementioned peptide; (iv) administration of antisense nucleic acid and nucleic acids that are "dysfunctional" (*i.e.*, due to a heterologous insertion within the coding sequences of coding sequences to an aforementioned peptide) that are utilized to "knockout" endogenous function of an aforementioned peptide by homologous recombination (*see, e.g.*, Capecchi, 1989. *Science* 244: 1288-1292); or (v) modulators (*i.e.*, inhibitors, agonists and antagonists, including additional peptide mimetic of the invention or antibodies specific to a peptide of the invention) that alter the interaction between an aforementioned peptide and its binding partner.

Diseases and disorders that are characterized by decreased (relative to a subject not suffering from the disease or disorder) levels or biological activity may be treated with Therapeutics that increase (*i.e.*, are agonists to) activity. Therapeutics that upregulate activity may be administered in a therapeutic or prophylactic manner. Therapeutics that may be utilized include, but are not limited to, an aforementioned peptide, or analogs, derivatives, fragments or homologs thereof; or an agonist that increases bioavailability.

Increased or decreased levels can be readily detected by quantifying peptide and/or RNA, by obtaining a patient tissue sample (*e.g.*, from biopsy tissue) and assaying it *in vitro* for RNA or peptide levels, structure and/or activity of the expressed peptides (or mRNAs of an aforementioned peptide). Methods that are well-known within the art include, but are not limited to, immunoassays (*e.g.*, by Western blot analysis, immunoprecipitation followed by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis, immunocytochemistry, *etc.*) and/or hybridization assays to detect expression of mRNAs (*e.g.*, Northern assays, dot blots, *in situ* hybridization, and the like).

Prophylactic Methods

In one aspect, the invention provides a method for preventing, in a subject, a disease or condition associated with an aberrant NOVX expression or activity, by administering to

the subject an agent that modulates NOVX expression or at least one NOVX activity.

Subjects at risk for a disease that is caused or contributed to by aberrant NOVX expression or activity can be identified by, for example, any or a combination of diagnostic or prognostic assays as described herein. Administration of a prophylactic agent can occur prior to the
5 manifestation of symptoms characteristic of the NOVX aberrancy, such that a disease or disorder is prevented or, alternatively, delayed in its progression. Depending upon the type of NOVX aberrancy, for example, a NOVX agonist or NOVX antagonist agent can be used for treating the subject. The appropriate agent can be determined based on screening assays described herein. The prophylactic methods of the invention are further discussed in the
10 following subsections.

Therapeutic Methods

Another aspect of the invention pertains to methods of modulating NOVX expression or activity for therapeutic purposes. The modulatory method of the invention involves contacting a cell with an agent that modulates one or more of the activities of NOVX protein
15 activity associated with the cell. An agent that modulates NOVX protein activity can be an agent as described herein, such as a nucleic acid or a protein, a naturally-occurring cognate ligand of a NOVX protein, a peptide, a NOVX peptidomimetic, or other small molecule. In one embodiment, the agent stimulates one or more NOVX protein activity. Examples of such stimulatory agents include active NOVX protein and a nucleic acid molecule encoding
20 NOVX that has been introduced into the cell. In another embodiment, the agent inhibits one or more NOVX protein activity. Examples of such inhibitory agents include antisense NOVX nucleic acid molecules and anti-NOVX antibodies. These modulatory methods can be performed *in vitro* (e.g., by culturing the cell with the agent) or, alternatively, *in vivo* (e.g., by administering the agent to a subject). As such, the invention provides methods of treating
25 an individual afflicted with a disease or disorder characterized by aberrant expression or activity of a NOVX protein or nucleic acid molecule. In one embodiment, the method involves administering an agent (e.g., an agent identified by a screening assay described herein), or combination of agents that modulates (e.g., up-regulates or down-regulates) NOVX expression or activity. In another embodiment, the method involves administering a
30 NOVX protein or nucleic acid molecule as therapy to compensate for reduced or aberrant NOVX expression or activity.

Stimulation of NOVX activity is desirable *in situations* in which NOVX is abnormally downregulated and/or in which increased NOVX activity is likely to have a beneficial effect. One example of such a situation is where a subject has a disorder characterized by aberrant cell proliferation and/or differentiation (*e.g.*, cancer or immune associated disorders).

- 5 Another example of such a situation is where the subject has a gestational disease (*e.g.*, preclampsia).

Determination of the Biological Effect of the Therapeutic

- In various embodiments of the invention, suitable *in vitro* or *in vivo* assays are performed to determine the effect of a specific Therapeutic and whether its administration is indicated for treatment of the affected tissue.

- In various specific embodiments, *in vitro* assays may be performed with representative cells of the type(s) involved in the patient's disorder, to determine if a given Therapeutic exerts the desired effect upon the cell type(s). Compounds for use in therapy may be tested in suitable animal model systems including, but not limited to rats, mice, chicken, cows, monkeys, rabbits, and the like, prior to testing in human subjects. Similarly, for *in vivo* testing, any of the animal model system known in the art may be used prior to administration to human subjects.

Prophylactic and Therapeutic Uses of the Compositions of the Invention

- The NOVX nucleic acids and proteins of the invention are useful in potential prophylactic and therapeutic applications implicated in a variety of disorders. The disorders include but are not limited to, *e.g.*, those diseases, disorders and conditions listed above, and more particularly include those diseases, disorders, or conditions associated with homologs of a NOVX protein, such as those summarized in Table A.

- As an example, a cDNA encoding the NOVX protein of the invention may be useful in gene therapy, and the protein may be useful when administered to a subject in need thereof. By way of non-limiting example, the compositions of the invention will have efficacy for treatment of patients suffering from diseases, disorders, conditions and the like, including but not limited to those listed herein.

- Both the novel nucleic acid encoding the NOVX protein, and the NOVX protein of the invention, or fragments thereof, may also be useful in diagnostic applications, wherein the presence or amount of the nucleic acid or the protein are to be assessed. A further use could be as an anti-bacterial molecule (*i.e.*, some peptides have been found to possess anti-bacterial

properties). These materials are further useful in the generation of antibodies, which immunospecifically-bind to the novel substances of the invention for use in therapeutic or diagnostic methods.

The invention will be further described in the following examples, which do not limit the scope of the invention described in the claims.

EXAMPLES

Example A: Polynucleotide and Polypeptide Sequences, and Homology Data

The NOV1 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 1A.

Table 1A. NOV1 Sequence Analysis			
	SEQ ID NO: 1	459 bp	
NOV1a, CG101036-01 DNA Sequence	GGTCACTTCTTGGTCCTGGTCCAGGGCCTGTACTGACCACCTCCACGTGCCACTGGGGCTGTA AGGAGGAATGGCGGCCGTGGGCAGCCTGCTTGGCCGGCTGAGGCAGAGCACCGTGAAGGCCAC CGGACCTGCACTCCGCCGCTGCACACATCCTCCTGGCGAGCTGACAGCAGCAGGGCCTCACT CACTCGTGTGCACGCCAGGCTTATGCACGACTCTACCCCGTGTGCTGGTGAAGCAGGATGG CTCCACCATCCACATCCGCTACAGGGAGCCACGGCGCATGCTGGCGATGCCCATAGATCTGGA CACCTGTCTCCTGAGGAGCGCCGGCCAGGCTGCGGAAGCGTGGGGCTCAGCTCCAGTCGAG GAAGGAGTACGAGCAGGAGCTCAGTGATGACTTGCATGTGGAGCGCTACCGACAGGTCTGGAC CAGGACCAAGAAGTGACC		
	ORF Start: ATG at 71		ORF Stop: TGA at 455
	SEQ ID NO: 2	128 aa	MW at 15008.1kD
NOV1a, CG101036-01 Protein Sequence	MAAVGSLGLRLRQSTVKATGPALRLHTSSWRADSSRASLTRVHRQAYARLYPVLLVKQDGST IHIRYREPRRLAMPIDLTLSPERRARLRKGAQLQSRKEYEQELSDDLHVERYRQVWTRT KK		

Further analysis of the NOV1a protein yielded the following properties shown in Table 1B.

Table 1B. Protein Sequence Properties NOV1a	
PSort analysis:	0.5756 probability located in nucleus; 0.5070 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.2297 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV1a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 1C.

Table 1C. Geneseq Results for NOV1a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB97259	Novel human protein SEQ ID NO: 527 - Homo sapiens, 128 aa. [WO200222660-A2, 21-MAR-2002]	1..128 1..128	126/128 (98%) 126/128 (98%)	1e-66
AAB63419	Human breast cancer associated antigen protein sequence SEQ ID NO:781 - Homo sapiens, 130 aa. [WO200073801-A2, 07-DEC-2000]	1..128 3..130	126/128 (98%) 126/128 (98%)	1e-66
AAU29307	Human PRO polypeptide sequence #284 - Homo sapiens, 164 aa. [WO200168848-A2, 20-SEP-2001]	5..128 41..164	119/124 (95%) 119/124 (95%)	3e-62
ABB69193	Drosophila melanogaster polypeptide SEQ ID NO 34371 - Drosophila melanogaster, 107 aa. [WO200171042-A2, 27-SEP-2001]	35..119 18..100	37/85 (43%) 61/85 (71%)	1e-15
AAM96532	Human reproductive system related antigen SEQ ID NO: 5190 - Homo sapiens, 68 aa. [WO200155320-A2, 02-AUG-2001]	5..46 16..57	37/42 (88%) 37/42 (88%)	4e-13

In a BLAST search of public sequence databases, the NOV1a protein was found to have homology to the proteins shown in the BLASTP data in Table 1D.

5

Table 1D. Public BLASTP Results for NOV1a				
Protein Accession Number	Protein/Organism/Length	NOV1a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CZ83	2810038N09Rik protein - Mus musculus (Mouse), 127 aa.	7..128 6..127	89/122 (72%) 97/122 (78%)	2e-44
Q8R030	Similar to mitochondrial ribosomal protein L55 - Mus musculus (Mouse), 134 aa.	1..128 7..134	89/128 (69%) 100/128 (77%)	4e-44
Q9VE04	CG14283 protein (Putative transcription factor) (RH10246p) - Drosophila melanogaster (Fruit fly), 107 aa.	35..119 18..100	37/85 (43%) 61/85 (71%)	3e-15
Q9TYJ8	Y66H1A.3 protein - Caenorhabditis elegans, 150 aa.	32..124 51..141	31/93 (33%) 48/93 (51%)	5e-07

Q9UJF2	Ras GTPase-activating protein nGAP (RAS protein activator like 1) - Homo sapiens (Human), 1139 aa.	69..128 938..998	20/61 (32%) 32/61 (51%)	0.19
--------	--	---------------------	----------------------------	------

Pfam analysis predicts that the NOV1a protein contains the domains shown in the Table 1E.

Table 1E. Domain Analysis of NOV1a			
Pfam Domain	NOV1a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 2.

The NOV2 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 2A.

Table 2A. NOV2 Sequence Analysis			
	SEQ ID NO: 3	1067 bp	
NOV2a, CG101055-01 DNA Sequence	CGGCTGCGCTTCTGCTGGAAACGCTTGCTGGCGCCTGTCACCGGTTCCCTCCATTTTGAAAGG GAAAAAGGCTCTCCACCCATTCCCCTGCCCTAGGAGCTGGAGCCGAGGAGCCGCGCTCA TGGCGTTCAGCCCGTGGCAGATCCTGTCCCGTGCAGTGGGCGAAATGGACGTGGTCTGCGG TACGCGCGGGGCGCGGCGAGGACGAGGCTGGCGGGCCGAGGGCGACCCGAGGAGGAGG ATTCGCAAGCCGAGACCAATCCTTGAGTTTCACTCGGATTCTGAAGTAATTTTGAGACTC CTGAAGCTGAAACCCGATCCGATCACCTTTCAAGGAGTCTGTGATCCATCACTCGGATTGG CAGGACCTGGGGCCAAAAGCCAAGAATCACAAGAAGCTGATGAACAGCTTGTAGCAGAAGTGG TTGAAAAATGTTTCACTAAGACTTGTCTAAACCTTCAGAAAATGAAGTGCCACAGCAGGCCA TTGACTCTCACTCAGTCAAGAATTTAGAGAAGAACCTGAACATGATTTTAGCAAAATTTCCA TCGTGAGGCCATTTTCAATAGAAACGAAGGATTCACGGATATCTCGGCAGTCTCTCGGAACAA AAGTAGCTCATGGCTGTGTAAGTGCAGTCTCAGGCAAGGCTCTGCCTTCCAGCCCGCAGACG CCCTCCAGGACGAGGCGATGACAGAAGGCGAGTGGGGGTACCCCTCAGGCCTCCGAGAAAG CTGATCTAAAAGCTGGCAACTCTGTCCAGAGCTTGTGCCAGCAGAAGAAGCAAGCTGAGAA AGCCCAAGCTGTCCCTTGAGGAAGAAAGCAATTGGAGGAGAGTTCTCAGACACCAACGCTG CTGTGGAGGGCACACCTCTCCCAAGGCATCCTATCACTTCAGTCTGAAGAGTTGGATGAGA ACACAAGTCTTTGCTAGGAGATGCCAGGTTCCAGAAGTCTCCCTGACCTTAAAGAACTC CCGGCACTCTCAGTAGTGACACCAACGACTCAGGGGTGGAGCTGGGGGCAGGTTGAATG		
	ORF Start: ATG at 126		ORF Stop: TGA at 1062
	SEQ ID NO: 4	312 aa	MW at 33130.0kD
NOV2a, CG101055-01 Protein Sequence	MAFSPWQILSPVQWAKWTWSAVRGAAGEDEAGGPEGDPPEEDSQAETKSLSFSSDSEGNFET PEAETPIRSPFKESCDPSLGLAGPGAQSKQESQEADEQLVAEVVEKCSSKTCSENEVPQQA IDSHSVKNFREPEHDFSKISIVRPFISIETKDSTDISAVLGTQVHGCVTAVSGKALPSSPPD ALQDEAMTEGSMGVTLASAEADLKAGNSCELVPSRRSKLRPKPVPLRKKAIIGGEFSDTNA AVEGTPLPKASYHFSPEELDENTSPLLGDARFQKSPDLKETPGTLSSDNDNSGVELGAR		
	SEQ ID NO: 5	2731 bp	
NOV2b, CG101055-02 DNA Sequence	CAGAGGTCTAGCAGCCGGGCGCGCGGGCCGCGGGGCTGAGGAGGCCACAGGACGGCGTCTT CCCGGCTAGTGAGCCCGCGCGGGGCGCGGCTGCGGCGCACCGTGAGGGGAGGAGGCCGAGG AGGACGCGGCGCGGCTGCGGCGGGAGGAAGCGCTCCACAGGGCCCGCAGCGCACTCGTT TAACCACATCCGCGCTCTGCTGGAAACGCTGCTGGCGCCTGTCACCGGTTCCCTCCATTTT GAAAGGGAAAAAGGCTCTCCCAACCATTCCTGCCCCTAGGAGCTGGAGCCGAGGAGCCG		

	<p>CGCTCATGGCGTT CAGCCCGTGGCAGATCCTGTCCCGTGCAGTGGGCGAAATGGACGTGGT CTGCGGTACGCGGCGGGGCGCGCGGCGAGGACGAGGCTGGCGGGCCGAGGGCGACCCCGAGG AGGAGGATTTCGAAGCCGAGACCAATCCTTGAGTTTCAGCTCGGATTCTGAAGGTAATTTTG AGACTCCTGAAGCTGAAACCCCGATCCGATCACCTTTCAAGGAGTCTGTGATCCATCACTCG GATTGGCAGGACCTGGGGCCAAAAGCCAAGAATCACAAGAAGCTGATGAACAGCTTGTAGCAG AAGTGGTTGAAAAATGTTTATCTAAGACTTGTCTAAACCTTCAGAAAAATGAAGTGCCACAGC AGGCCATTGACTCTCACTCAGTCAAGAATTCAGAGAAGAACCTGAACATGATTTTAGCAAAA TTTCCATCGTGAGGCCATTTTCAATAGAAAAGGATTCCACGGATATCTCGGCAGTCCCTCG GAACAAAAGCAGCTCATGGCTGTGTAAGTGCAGTCTCAGGCAAGGCTCTGCCTTCAGCCCCG CAGACGCCCTCCAGGACGAGGCGATGACAGAAGGCAGCATGGGGGTACCCCTCGAGGCCCTCG CAGAAGCTGATCTAAAGCTGGCAACTCCTGTCCAGAGCTTGTGCCCAGCAGAAGAAGCAAGC TGAGAAAGCCCAAGCTGTCCCTTGAGGAAGAAAGCAATTGGAGGAGAGTTCTCAGACACCA ACGCTGTGTGGAGGGCACACTTCCCCAAGGCATCCTATCACTTCAGTCTGAAGAGTTGG ATGAGAACAACAAGTCTTTGCTAGGAGATGCCAGGTTCCAGAAGTCTCCCCCTGACATTAAG AAACTCCCGCACTCTCAGTAGTGACACCAACGACTCAGGGGTTGAGCTGGGGGAGGAGTCGA GGAGCTCACCTCTCAAGCTTGAGTTTGATTTACAGAAGATACAGGAAACATAGAGGCCAGGA AAGCCCTTCCAAGGAAGCTTGGCAGGAAACTGGGTAGCACACTGACTCCCAAGATACAAAAAG ATGGCATCAGTAAGTCAGCAGGTTTGAACAGCCTACAGACCCAGTGGCAGGAGGCCCTC TCTCCCAACATCTTCCAAGCCAGATCCTAGTCAGTGGGAGAGCCCCAGCTTCAACCCCTTTG GGAGCCACTCTGTTCTGCAGAACTCCCCACCCTCTCTCTGAGGGCTCTTACCCTTTGACC CAGATAACTTTGACGAATCCATGGATCCCTTTAAACCAACTACGACCTTAACAAGCAGTGACT TTTGTCTCCCACTGGTAATCACGTTAATGAAATCTTAGAATCACCAAGAAGGCAAGTCGC GTTTAATAACGAGTGGCTGTAAGGTGAAGAAGCATGAAACTCAGTCTCTCGCCCTGGATGCAT GTTCTCGGATGAAGGGGAGTGTCTCCAGATTTTCAAGATTTCTAATAGGGATGGCCATG CTACTGATGAGGAGAACTGGCATCCACGTATGTGGTCAGAATCAGTGGTCCGAGGTGA AAGGTGAGCCAGAGGAAGACCTGGAGTACTTTGAATGTTCCAATGTTCTGTGTCTACCATAA ATCATGCGTTTTCATCCTCAGAAGCAGGCATAGAGAAGGAGAGCTGCCAGAAGATGGAAGAAG ACGGGTCCACTGTGCTTGGGCTGCTGGAGTCTCTGCAGAGAAGGCCCTGTGTGGTGTCTCT GTGGAGGTGAGAGCCCCCTGGATGGATCTGCCTCAGCGAATCAGACAGACAGCCGTGCTCA CCTTAATAAGAGAAGAGATAATTACTAAAGAGATTGAAGCAATGAATGGAAGAAGAAATACG AAGAGACCCGGCAAGAAGTTTGGAGATGAGGAAATGTAGCTGAATATGAAAAGACTATTG CTCAAATGATTGAAGATGAACAAAGGACAAGTATGACCTCTCAGAAGAGCTTCCAGCACTGA CCATGGAGAAGGAACAGGCCCTGGCTGACCTTAACCTCTGTGGAAGGTCCCTTTCTGATCTCT TCAGGAGATATGAGAACCTGAAAGGTGTTCTGGAAGGGTTCAAGAAGATGAAGAAGCCTGA AGAAATGTGCTCAGGATTACTTAGCCAGAGTTAAACAAGAGGAGCAGCGATACCAGGCCCTGA AAATCCACGCAGAAGAGAACTGGACAAGTAAGAGCTTGTAATGTTGAATTTCACTCTTCAT GATGTTGTGGGAAGATTGAGAGAGGAAAACAAATCACTGTTTCGCAACTCCAGGTTGTATTT TTATGTGTGTGTTTATTTCACTTTTAAACCTTTTCCCATTTGTAATAAAAAAAAAAAAAAAA AAAAAAAAAAAAACCCAAAAA</p>
	<p>ORF Start: ATG at 321</p> <p>ORF Stop: TAA at 2550</p>
	<p>SEQ ID NO: 6</p> <p>743 aa</p> <p>MW at 80863.4kD</p>
NOV2b, CG101055-02 Protein Sequence	<p>MAFSPWQILSPVQWAKWTWSAVRGGAAAGEDEAGGPEGDPEDDSQAETKSLFSFSSDSEGNFET PEAETPIRSPFKESCDPSLGLAGPAKQSQESQEADELVAEVVEKCSKTSKPSENEVPQQA IDSHSVKNFREPEHDFSKISIVRPFISIETKDSTDISAVLGTAAHGCVTAVSGKALPSSPPD ALQDEAMTEGSMGVTLEASAEADLKAGNSCELVPSRRSKLRKPKVPLRKAIGGEFSDTNA AVEGTPLPKASYHFSPEELDENTSPLLGDARFQKSPDIIKETPGTLSSDNDSGVELGEESRS SPLKLEFDFTEDTGNIEARKALPRKLGRKLGLSTLTPKIQKDGISKSAGLEQPTDPVARDGPLS QTSSKPDPSQWESPSFNPFGSHSVLQNSPPLSSEGSYHFDPDNFDSEMDPFPKPTTTLTSSDFC SPTGNHVNEILESPPKAKSRLITSGCKVKKHETQSLALDACSRLDEGAVISQISDISNRDGHAT DEEKLASTSCGQKSAGAEVKGEPEEDLEYFECSNVPVSTINHAFFSSSEAGIEKETCQKMEEDG STVLGLLESSAEKAPVSVSCGESPLDGICLSESDKTAVLTILIREEIIITKEIANEWKKKYEE TRQEVLEMRKIVAEYEKTI AQMIEDEQRTSMTSQSFQQLTMEKEQALADLNSVERSLSDLFR RYENLKGVLGFKKNEEALKKCAQDYLARVKQEEQRYQALKIHAEELDK</p>

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 2B.

Table 2B. Comparison of NOV2a against NOV2b.

Protein Sequence	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV2b	1..310 1..310	269/310 (86%) 270/310 (86%)

Further analysis of the NOV2a protein yielded the following properties shown in Table 2C.

Table 2C. Protein Sequence Properties NOV2a

PSort analysis:	0.8800 probability located in nucleus; 0.4612 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1582 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 21 and 22

5

A search of the NOV2a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 2D.

Table 2D. Geneseq Results for NOV2a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM38680	Human polypeptide SEQ ID NO 1825 - Homo sapiens, 1025 aa. [WO200153312-A1, 26-JUL-2001]	1..295 1..287	79/318 (24%) 118/318 (36%)	2e-08
AAM38679	Human polypeptide SEQ ID NO 1824 - Homo sapiens, 966 aa. [WO200153312-A1, 26-JUL-2001]	1..295 1..287	79/318 (24%) 118/318 (36%)	2e-08
AAM38678	Human polypeptide SEQ ID NO 1823 - Homo sapiens, 1013 aa. [WO200153312-A1, 26-JUL-2001]	1..295 1..287	79/318 (24%) 118/318 (36%)	2e-08
ABG22566	Novel human diagnostic protein #22557 - Homo sapiens, 637 aa. [WO200175067-A2, 11-OCT-2001]	1..96 56..138	33/96 (34%) 49/96 (50%)	1e-07
ABG16821	Novel human diagnostic protein #16812 - Homo sapiens, 124 aa. [WO200175067-A2, 11-OCT-2001]	1..73 56..115	29/73 (39%) 43/73 (58%)	2e-07

10

In a BLAST search of public sequence databases, the NOV2a protein was found to have homology to the proteins shown in the BLASTP data in Table 2E.

Table 2E. Public BLASTP Results for NOV2a				
Protein Accession Number	Protein/Organism/Length	NOV2a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75410	Transforming acidic coiled-coil-containing protein 1 - Homo sapiens (Human), 805 aa.	1..310 1..310	308/310 (99%) 309/310 (99%)	e-179
Q9PTG8	CPEB-associated factor Maskin - Xenopus laevis (African clawed frog), 931 aa.	215..308 260..362	34/106 (32%) 47/106 (44%)	0.015
Q8SY55	GH09355p - Drosophila melanogaster (Fruit fly), 1514 aa.	29..306 875..1136	66/284 (23%) 117/284 (40%)	0.034
Q8T4F6	SD08609p - Drosophila melanogaster (Fruit fly), 944 aa.	34..144 562..678	32/123 (26%) 56/123 (45%)	0.044
Q9VKQ6	CG6729 protein - Drosophila melanogaster (Fruit fly), 944 aa.	34..144 562..678	32/123 (26%) 56/123 (45%)	0.044

- 5 Pfam analysis predicts that the NOV2a protein contains the domains shown in the Table 2F.

Table 2F. Domain Analysis of NOV2a			
Pfam Domain	NOV2a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 3.

- 10 The NOV3 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 3A.

Table 3A. NOV3 Sequence Analysis			
	SEQ ID NO: 7	478 bp	
NOV3a, CG101973-01 DNA Sequence	CGGAAGAGGGGGTGAAGGCCAGAGGCTCGGGGGCTTCAAGACCGCTGTCTGGAGTCCCCCTTTC CAGGCCATGTGCGGGGCCACCTGGCTGCCCCGAAGCAGCCGAGCCCGCAGAGCCCCCTCAG GGGAGGGCGATCCCCCGCGGCACCCGCGGGCCACCAACCGGCCACGGAGCAGGGGCTCCCTGC AGACAGGGGGGCCTTCGCCCTGGAAGCCTGGACGCCGAGATAGACTTGCTGAGCAGCACGCT GGCCGAGCTGAATGGGGTTCGGGGTCATGCGTCACGGCGACCAGACCGACAGGCATATGAGCC		

	CCCGCCACCTCCTGCCTACCGCAGGGCTCCCTGAAGCCAAATCCAGCCTCGCCGCTCCCAGC GTCTCCCTATGGGGGCCCCACTCCAGCCTCTTACACTACCGCCAGCACCCCGGCTGGCCAGC CTTCCCCGTGCAAGTGAAGGTGGCACAGCCAGTGAGG		
	ORF Start: ATG at 70		ORF Stop: TGA at 310
	SEQ ID NO: 8	80 aa	MW at 8277.3kD
NOV3a, CG101973-01 Protein Sequence	MSGPTWLPPKQPEPARAPQGRAIPRGTPGPPPAHGAGAPCRQGGSPWKPGRDRRLAEQHAGR AEWGS GSCVTATRP TGI		
	SEQ ID NO: 9	1673 bp	
NOV3b, CG101973-02 DNA Sequence	CGGAAGAGGGGGTGAAGGCCAGAGGCTCGGGGCTTCAAGACCGTGTCTGGAGTCCCCCTTTC CAGGCCATGTGCGGGGCCACCTGGCTGCCCCGAAGCAGCGGAGCCCGCAGAGCCCTCAG GGGAGGGCGATCCCCGCGGCACCCGGGGGCCACCACGGGCCACGGAGCAGCACTCCAGCCC CAGCCAGGGTCAATTTTGCCTTCCATCTGAGCAGTGTACCAGGCCCCAGGGGGACCG GAGGATCGGGGGCCGCGTGGGTGGGTCCCAGGAGTACTCCAGCACACGAGGGGCTCCCT GCAGACAGGGGGGCTTCGCTTGAAGCCTGGACGCGGAGATAGACTTGCTGAGCAGCAGC CTGGCCGAGCTGAATGGGGTGGGGTCTGCGTACGCGGACAGACCGACAGGCATATGAG CCCCCGCCACCTCCTGCCTACCGCAGGGGCTCCCTGAAGCCAAATCCAGCCTCGCCGCTCCA GCGTCTCCCTATGGGGGCCCCACTCCAGCCTCTTACACTACCGCCAGCACCCCGGCTGGCCCA GCCTTCCCCGTGCAAGTGAAGGTGGCACAGCCAGTGAGGGGCTGCGGCCACCCAGGCGGGGA GCCTCTCAGGCCTCTGGGCCCCCTCCGGGCCCCACTTTCTCTCCAGGCCAGGTGAAGTC TGGGGGCTGGCTATAGGAGCCAGAGAGAGCCAGGGCCAGGGGCCAAAGAGGAAGCTGCTGGG GTCTCTGGCCCTGCAGGAAGAGGAAGAGGAGGCGAGCAGGGGCCAGGTGCCCTGAGCCAG CCTCCAGGAGTGAAGTGGATAGGCTGACGAAGAAGCTGTTTCACGACATGAACACCCGCC AGCGGGGAGTACTTTGGCCAGTGTGGTGGCTGCGGAGAAGATGTGGTTGGGATGGGGCTGGG GTTGTGGCCCTTGATCGGCTCTTTCACGTGGGCTGCTTTGTATGTTCTACATGCCGGGCCAG CTTCGCGGCCAGCATTTCTACGCGCTGGAGAGGAGGGCATATTGCAGGGGCTGTACTGTGGCC ACCTGGAGAAATGTGCCAGTGTCTCCAGCCCATCTGGACCGGATCCTGCGGGCTATGGGG AAGGCCTACCACTGGCTGCTTCACTGCGTGGTGTGTACCGCGGCTCAGCGGCATCCCC TTCACAGTGGATGCTACGAGCCAGATCCACTGCATTGAGGACTTTCACAGGAAGTTTGCCCCA AGATGCTCAGTGTGCGGTGGGGCCATAATGCTGAGCCAGGTGAGGAGGAGACTGTGAGAATT GTTGCTCTGGATCGAAGTTTTCACATTGGCTGTTACAAGTGCAGGAGTGTGGGCTGCTGCTC TCCTCTGAGGGCGAGTGTGAGGGCTGTACCCGCTGGATGGGCACATCTGTGCAAGGCCTGC AGCGCCTGGCGCATCCAGGAGCTCTCAGCCACCGTCACCACTGACTGCTGAGTCTTCTAGAA GTACCTGCTGGGTCTCAGTTCCAGTTCATCTTTGATTGATCACTCTCCTTGACATCCAC CTGTATGACTTTGTACCAAATGCTGTCTTCTTCTCTCAATCAAGAAATAATAATCCCTCG AGTTTACAAAACAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 70		ORF Stop: TGA at 1498
	SEQ ID NO: 10	476 aa	MW at 50287.4kD
NOV3b, CG101973-02 Protein Sequence	MSGPTWLPPKQPEPARAPQGRAIPRGTPGPPPAHGAALQPHPRVNFCLPSEQCYQAPGGPED RGPWVSGSHVLTQGLPADRGGLRPGSLDAEIDLSSTLAELNGRGHASRRPDRQAYEPP PPPAYRTGSLKPNPASPLPASYPGPTPASYTASTPAGPAFPVQVKVAQPVRCGPPRRGAS QASGPLPGPHFPLPGRGEVWPGYRSQREPGPAKEEAAGVSGPAGRGRGGEHGPQVPLSQPP EDELDRLTKKLVHDMNHPPSGEYFGQCGGCGEDVVDGAGVVALDRV FHVGCFCVSTCRAQLR GQHFYAVERRAYCEGYVATLEKCATCSQPILDRI LRAMGKAYHPGFTCVVCHRGLDGIPT VDATSIHICIEDFHRKFAPRCVCGGAIMPEPGQETVRIVALDRSFHIGCYKCEECGLLLSS EGECQGCYPLDGHILCKACSAWRIQELSATVTDDC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 3B.

Table 3B. Comparison of NOV3a against NOV3b.		
Protein Sequence	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV3b	1..19	19/19 (100%)
	1..19	19/19 (100%)

Further analysis of the NOV3a protein yielded the following properties shown in Table 3C.

Table 3C. Protein Sequence Properties NOV3a	
PSort analysis:	0.8486 probability located in lysosome (lumen); 0.4500 probability located in cytoplasm; 0.2583 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV3a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 3D.

Table 3D. Geneseq Results for NOV3a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM52307	Human TRIP6 - Homo sapiens, 476 aa. [WO200171356-A2, 27-SEP-2001]	1..36 1..36	36/36 (100%) 36/36 (100%)	7e-17
AAM52308	Murine TRIP6 - Mus musculus, 480 aa. [WO200171356-A2, 27-SEP-2001]	1..36 1..36	29/36 (80%) 32/36 (88%)	1e-12
AAR79480	Rat type II collagen - Rattus sp, 1442 aa. [WO9522611-A2, 24-AUG-1995]	1..79 952..1029	29/80 (36%) 34/80 (42%)	0.011
AAE16477	Human collagen alpha1 (II) protein - Homo sapiens, 1418 aa. [US6323314-B1, 27-NOV-2001]	1..79 928..1005	28/80 (35%) 34/80 (42%)	0.024
ABB09627	Amino acid sequence of human collagen type II alpha1 - Homo sapiens, 1418 aa. [US6342361-B1, 29-JAN-2002]	1..79 928..1005	28/80 (35%) 34/80 (42%)	0.024

10

In a BLAST search of public sequence databases, the NOV3a protein was found to have homology to the proteins shown in the BLASTP data in Table 3E.

Table 3E. Public BLASTP Results for NOV3a				
Protein Accession Number	Protein/Organism/Length	NOV3a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAC94501	Sequence 15 from Patent WO0171356 - Homo sapiens (Human), 476 aa.	1..36 1..36	36/36 (100%) 36/36 (100%)	2e-16
Q9BUE5	Similar to thyroid hormone receptor interactor 6 - Homo sapiens (Human), 474 aa.	1..36 1..36	36/36 (100%) 36/36 (100%)	2e-16
Q9BXP3	Thyroid receptor interacting protein 6 - Homo sapiens (Human), 476 aa.	1..36 1..36	36/36 (100%) 36/36 (100%)	2e-16
Q15654	Thyroid receptor interacting protein 6 (TRIP6) (OPA-interacting protein 1) (Zyxin related protein 1) (ZRP-1) - Homo sapiens (Human), 476 aa.	1..36 1..36	36/36 (100%) 36/36 (100%)	2e-16
Q9Z1Y4	Zyxin related protein-1 (Thyroid hormone receptor interactor 6) (TRIP6) - Mus musculus (Mouse), 480 aa.	1..36 1..36	29/36 (80%) 32/36 (88%)	3e-12

PFam analysis predicts that the NOV3a protein contains the domains shown in the Table 3F.

Table 3F. Domain Analysis of NOV3a			
Pfam Domain	NOV3a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 4.

The NOV4 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 4A.

Table 4A. NOV4 Sequence Analysis			
	SEQ ID NO: 11	423 bp	
NOV4a, CG102244-01 DNA Sequence	GGGTCCTGGGGCCCGAGCGGGTGGCCCGCGGGCCCTCGGGCTGCGTGGGGAGGGGGCTTC CGCCCCCTGTTGTCATTGCTCCTGCAGCCTTTTCGCTGGGACTGCGCGACACCGCCCCCGACC GGGTGCCCGCTGTGTGCCAGGCCGGGTGCTGGGCACGGTCCCGCGAGTGCCCTATAAGGACTG CCAGGCAATAATGAAGGTCTTTTACTGAAGGATGCGAAGGAAGATGACTGTGGCCAGGATCC GTATATCAGGGAATTAGGATTATATGGACTTGAAGCCACTTTGATCCCTGTTTATCGTTTGA GTTTTTGTCTCTCCAGTTTCTCTGAGAAGTCTGGGAAAGGTCTCTGAAAGAAAAATGGAAT GCCAAGTCAGTGTATGTGTTGGAAATGCTACTGCTTCTCTAGTG		
	ORF Start: ATG at 200		ORF Stop: TGA at 362

	SEQ ID NO: 12	54 aa	MW at 5975.9kD
NOV4a, CG102244-01 Protein Sequence	MKVLLKDAKEDDCGQDPYIRELGlyGLEATLIPVLSFEFLSLPSFSEKSGKGL		
	SEQ ID NO: 13	1086 bp	
NOV4b, CG102244-02 DNA Sequence	CCACAGAGGGCAGTCACGTGCCCGCTGTGTGCCAGGCCGGGTGCTGGGCACGGTCCCGCGAGT GCCCTATAAGGACTGCCAGGCAATAATGAAGGTTCTTTACTGAAGGATCGGAAGGAAGATGA CTGTGGCCAGGATCCGTATATCAGGGAATTAGGATTATATGGACTTGAAGCCACTTTGATCCC TGTTTTATCGTTTGAGTTTTTGTCTCTTCCAGTTTCTCTGAGAAGCTTTCTCATCCTGAAGA TTACGGGGGACTCATTTTTACCAGCCCCAGAGCAGTGGAAGCAGCAGAGTTATGTTTGAGCA AAACAATAAACTGAAGTCTGGGAAAGGTCTCTGAAAGAAAATGGAATGCCAAGTCAGTGTA TGTGGTTGGAAATGCTACTGCTTCTCTAGTGAGTAAAATTGGCCTGAATACAGAAGGAGAAAC CTGTGGAATGCAGAAAAGCTTGCAGAATATATTGTTCCAGGGAGTCTCAGCACTGCCTCT TCTATTTCCCTGTGGAACCTCAAAGAGAAATCCTGCCAAAAGCGCTCAAGGACAAAGGGAT TGCCATGGAAGCATAACTGTGTATCAGACAGTTGCACACCCAGGAATCCAAGGGAACCTGAA CAGCTACTATTCCCAGCAGGGGTTCCAGCCAGCATCACATTTTTAGTCCCTCTGGCCTCAC ATACAGTCTCAAGCACATTCAGGAGTTATCTGGTGACAATATCGATCAAATTAAGTTTGACAGC CATCGGCCCACTACGGCTCGCGCGCTGGCCGCCCAGGGCCTTCTGTAAGCTGCACTGCAGA GAGCCCCACGCCACAAGCCCTGGCCACTGGCATCAGGAAGGCTCTCCAGCCCCATGGCTGCTG CTGAGTCAGCCACCTAGCGCTGGCCCCATGCAGCCTCCCTGGGCTGGGCTGGCTCTGGATGGA GCCAGGCATCGGCAAGGGCTCTCGGGAGCTGCTGCCGTCAAGCTCCTGCCTCAAGCCTGAGTG GAAGCACCTGAGGACCGGGATCGGGACCTGACCTGGGGCTGGCCTCAGGCCACGTGCACGT GACTGCCCTCTGTGG		
	ORF Start: ATG at 89		ORF Stop: TGA at 884
	SEQ ID NO: 14	265 aa	MW at 28626.3kD
NOV4b, CG102244-02 Protein Sequence	MKVLLKDAKEDDCGQDPYIRELGlyGLEATLIPVLSFEFLSLPSFSEKLSHPEDYGGIIFTS PRAVEAAELCLEQNNKTEVWERSLKEKWNKSVYVVG NATASLVSKIGLNTGEETCGNAEKLA EYICRESSALPLLPFCGNLKRILPKALKDKGIAMESITVYQTVHPGIQGNLNSVYSQQGV PASITFFSPSGLTYSCLKHIQELSGDNIDQIKFAAIGPTTARALAAQGLPVSCTAESPFPQALA TGIRKALQPHGCC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 4B.

Table 4B. Comparison of NOV4a against NOV4b.		
Protein Sequence	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV4b	1..35 1..35	35/35 (100%) 35/35 (100%)

5

Further analysis of the NOV4a protein yielded the following properties shown in Table 4C.

Table 4C. Protein Sequence Properties NOV4a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0480 probability located in microbody (peroxisome)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV4a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 4D.

5

Table 4D. Geneseq Results for NOV4a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

In a BLAST search of public sequence databases, the NOV4a protein was found to have homology to the proteins shown in the BLASTP data in Table 4E.

Table 4E. Public BLASTP Results for NOV4a				
Protein Accession Number	Protein/Organism/Length	NOV4a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P10746	Uroporphyrinogen-III synthase (EC 4.2.1.75) (UROS) (Uroporphyrinogen- III cosynthetase) (Hydroxymethylbilane hydrolyase [cyclizing]) (UROIIIS) - Homo sapiens (Human), 265 aa.	1..49 1..49	49/49 (100%) 49/49 (100%)	2e-21
P51163	Uroporphyrinogen-III synthase (EC 4.2.1.75) (UROS) (Uroporphyrinogen- III cosynthetase) (Hydroxymethylbilane hydrolyase [cyclizing]) (UROIIIS) - Mus musculus (Mouse), 265 aa.	1..49 1..49	42/49 (85%) 44/49 (89%)	4e-16
Q9CW78	Uroporphyrinogen III synthase - Mus musculus (Mouse), 27 aa (fragment).	1..27 1..27	22/27 (81%) 23/27 (84%)	9e-05
Q9JLU5	Uroporphyrinogen III synthase (EC 4.2.1.75) - Mus musculus (Mouse), 21 aa (fragment).	1..21 1..21	18/21 (85%) 19/21 (89%)	0.005

10

PFam analysis predicts that the NOV4a protein contains the domains shown in the Table 4F.

Table 4F. Domain Analysis of NOV4a			
Pfam Domain	NOV4a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 5.

The NOV5 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 5A.

5

Table 5A. NOV5 Sequence Analysis			
	SEQ ID NO: 15	1073 bp	
NOV5a, CG102713-01 DNA Sequence	GACACCCCTCTCTGTGACTCAGTCTCTGAGCGTTTTAATACGATGGTGTCCCGCGGGATCAA ACTTCAGCGTCACAGCTGAGGACTGGCTTCGTGGTCCCTGATGGGAGAGCATGAACAGGTGGT ATGTGAAGCCCTTGGAGACCAGCTCTTCCAAAGTCAAAGCCAAGACCATTGTGATGATTCTCG ACTCCCAGAAGCTCCTGCCATGTGAACCTGAGTCACTCAAGAGCCAGTTACAGGCCAGACCA AGGCTTTCGAGTTCCTGAACCACTCAGTGACCATGTTGGAGAAGGAGAGCTGCTTGCAGCAA TCAAGATTCAGCAGCTTGAAGAGGTGCTGAGCCCCACAGGCCGCCAGGGAGAGAAGGAGGAGC ACAAGTGGGGCATGGAGCAGGGCCGGCAGGAGCTGTATGGGGCCCTGACCAAGGCCTTCAGG GGCTGGAGAAGACCCTGCGTGACAGTGAGGAGATGCAGCGGGCCCGCACCCTCGCTGCCTGC AGCTGCTGCGCCAGGAGATCCGGGACAGCAAGAAGTTCCTGTGGGAGGAGCTGGAAGTGGTGC GGGAGGAGGTGACCTTCATCTATCAGAAGCTCCAAGCGCAGGAGGATGAGATCTCAGAGAACT TGGTGAACATTTCAGAAAATGCAGAAAACGAGGTGAAATGCCGAAAATCCTGACCAAGATGA AGCAGCAGGGTCATGAGACAGCCGCTGTCCGGAGACTGAAGAGATACCGCAGGGAGCCAGTG GCTGCTGGAAGGATGACCTCCAGAAGGAACTGAGTGATATATGGTCTGCTGTGCACGTGCTGC AGAACTCCATAGACAGCCTCACTTTGTGCTCGGGGGCCTGTCCCAAGGCCTCGAGCCTAAGAG GCCACAAGGGGCACCAAGTGCCTGAGCCCTCACTCCCTCCTGGGACTCTGACTCCGACTCTG ACCAGGACCTCTCCAGCCACCTTTCAGCAAGAGTGGCCGCTCCTTCCCACCGCTTGAGCAG CCGGGACTGCTCTCCCTGAAGACCCCTCCAGAGAGAAAATAAAGTAGCCAGACCCTCTCTA AA		
	ORF Start: ATG at 114		ORF Stop: TGA at 1002
	SEQ ID NO: 16	296 aa	MW at 33734.1kD
NOV5a, CG102713-01 Protein Sequence	MNRWYVKPLETSSSKVKAKTIVMILDSQKLLRCELESLSQLQAQTKAFELNHSVTMLEKES CLQQIKIQOLEEVLSPTRGRQGEKEEHKWGMEQGRQELYGALTQGLGLEKTLRDSSEMQRRART TRCLQLLAQEIIRDSKKFLWEELELVREEVTFIYQKLQAQDEISENLVNIQKMQTKQVCKRKI LTKMKQGHETAACPETEIPQASGCWKDDLQKELSDIWSAVHVLQNSIDSLTLCSGACPKA SSLRGHKHQCLSPPLPSWSDSDSDQDLSPFPFSKSGRSFPPA		
	SEQ ID NO: 17	1382 bp	
NOV5b, CG102713-02 DNA Sequence	TTCAGCGTCACAGCTGAGGACTGGCTTCGTGGTCCCTGATGGGAGAGCATGAACAGGACCCTC TTTTGGCCGGCTACCCCGGGACCCTGACTACTCTGTGCTCCTGCTCTACTACCTCCCTCAA GGAAGCCCTACCCCTAGGCTCTTCTCGGCAAGGCTCTGGAGCGTACAGCTCACTGGTCCA GGACTCCAGAGCCAGAGACCTTGGGATGCCCTGCTTCTGGGGACACAGTGAGGACTGCAGACT GCAGGCCAGGGTGGGGCTCAGGGCCTTCGCCACATGAGGCTGCCCCCTCCCCAGTCCAGACC TGCAGAAGCAGTGCTGTAATGACCAGGACATTTGAAGAGGCATCACAACGTAGCTAAGGTCA CCCCCTCTTCTGCCCCACAGCAAGACCATTGTGATGATTCCCGACTCCCAAGAGCTCCTGC GATGTGAAGTTCAGTCACTCAAGAGCCAGTTACAGGCCAGACCAAGGCTTTCGAGTTCCTGA ACCACTCAGTGACCATGTTGGAGAAGGAGAGCTGCTTGCAGCAATCAAGATTACGACGTTG AAGAGGTGCTGAGCCCCACAGGCCGCCAGGGAGAGAAGGAGGAGACAAGTGGGGCATGGAGC AGGGCCGGCAGGAGCTGTATGGGGCCCTGACCAAGGCCTTCAGGGGCTGGAGAAGACCCTGC GTGACAGTGAGGAGATGCAGCGGGCCCGCACCCTCGCTGCCTGCAGTGTCTGGCCAGGAGA TCCGGGACAGCAAGAAGTTCCTGTGGGAGGAGTGGAAGTGGTGCAGGAGGAGTACCTTCA TCTATCAGAAGCTCCGTGAGCAGGAGGATGAGATCTCAGAGAACTTGGTGAACATTTCAGAAA TGCAGAAAACGAGGTGAAATGCCGCAAGTGCTGACCAAGATGAAGCAGCAGGGTCATGAGA CAGCCGCTGTCCGGAGACTGAAGAGATACCGCAGGGAGCCAGTGGCTGCTGGAAGGATGACC		

	TCCAGAAGGAACTGAGTGATATATGGTCTGCTGTGCACGTGCTGCAGAACTCCATAGACAGCCTCACTTTGTGCTCGGGGGCCTGTCCCAAGGCCTCGAGCCTAAGAGGCCACAAGGGGCACCACTGCCTGAGCCCTCCACTCCCCTCCTGGGACTCTGACTCCGACTGTGACCAGGACCTCTCCAGCCACCTTTAGCAAGAGCGGCGGCTCCTTCCCACCCGGTGAGATCCTCCCCAGTCCCCCCTCACCCATTTCCCTCCTGACCTGCCCTTGACTCCACAGCTAGACCCTCAGAACAGGCCCGAGATCCTGATCTGGGCATGCGGTCCCTGACCTGCAGCCCCGGAGTCCCCTGGACCTGGCCAG		
	ORF Start: ATG at 39		ORF Stop: TGA at 1287
	SEQ ID NO: 18	416 aa	MW at 46300.7kD
NOV5b, CG102713-02 Protein Sequence	MGEHEQDPLLAGLPRDPDYSVSLYSPPSRKPSPLGSSSDKALERTAHWSRTPPETLGCPASGDTVRTADCRPGWGSPPHEAAPSPPDLQKCCNDQDILKRHHNVAKVTPSSLPTAKTIVMIPDSQKLLRCELESLSQLQAQTKAFELNHSVTMLEKESCLQIKIQGLEEVLSPTRQGEKEEHKWMGEQGRQELYGALTQGLQGLEKTLRDSEEMQARTRCLQLLAQEIRDSKKFLWEEELVREEVTFIYQKLREQEDEISENLVNIQKMQKTQVKCRKVLTKMKQGHETAACPETEIPQASGCNKDDLQKELSDIWSAVHVLQNSIDSLTLCGACPKASSLRGHKGHCSPPLPSWSDSDCDQDLSQPPFSKSGRSFPPGADPPQSPPPPISLLTCP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 5B.

Table 5B. Comparison of NOV5a against NOV5b.		
Protein Sequence	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV5b	11..293 114..396	245/283 (86%) 247/283 (86%)

5

Further analysis of the NOV5a protein yielded the following properties shown in Table 5C.

Table 5C. Protein Sequence Properties NOV5a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3600 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV5a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 5D.

Table 5D. Geneseq Results for NOV5a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AA Y60569	Human normal bladder tissue EST encoded protein 241 - Homo sapiens, 227 aa. [DE19818620-A1, 28-OCT-1999]	1..211 4..214	210/211 (99%) 210/211 (99%)	e-116
AAG03416	Human secreted protein, SEQ ID NO: 7497 - Homo sapiens, 87 aa. [EP1033401-A2, 06-SEP-2000]	6..86 7..87	79/81 (97%) 79/81 (97%)	2e-36
AAM35994	Peptide #10031 encoded by probe for measuring placental gene expression - Homo sapiens, 40 aa. [WO200157272-A2, 09-AUG-2001]	189..228 1..40	40/40 (100%) 40/40 (100%)	4e-17
AAM20726	Peptide #7160 encoded by probe for measuring cervical gene expression - Homo sapiens, 40 aa. [WO200157278-A2, 09-AUG-2001]	189..228 1..40	40/40 (100%) 40/40 (100%)	4e-17
AAM75883	Human bone marrow expressed probe encoded protein SEQ ID NO: 36189 - Homo sapiens, 40 aa. [WO200157276-A2, 09-AUG-2001]	189..228 1..40	40/40 (100%) 40/40 (100%)	4e-17

In a BLAST search of public sequence databases, the NOV5a protein was found to have homology to the proteins shown in the BLASTP data in Table 5E.

Table 5E. Public BLASTP Results for NOV5a				
Protein Accession Number	Protein/Organism/Length	NOV5a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CXZ5	2510048L02Rik protein - Mus musculus (Mouse), 403 aa.	7..296 125..403	220/290 (75%) 242/290 (82%)	e-117
Q9D4W5	2510048L02Rik protein - Mus musculus (Mouse), 267 aa.	7..139 133..262	106/133 (79%) 116/133 (86%)	1e-51
P30427	Plectin 1 (PLTN) (PCN) - Rattus norvegicus (Rat), 4687 aa.	6..211 2592..2783	55/210 (26%) 101/210 (47%)	2e-07
Q9CW93	0610037D15Rik protein - Mus musculus (Mouse), 322 aa (fragment).	34..180 81..223	47/150 (31%) 74/150 (49%)	3e-06
Q9JI55	Plectin 1 (PLTN) (PCN) (300-kDa intermediate filament-associated protein) (IFAP300) - Cricetulus griseus (Chinese hamster), 4473 aa (fragment).	6..213 2378..2571	54/212 (25%) 100/212 (46%)	7e-06

5

PFam analysis predicts that the NOV5a protein contains the domains shown in the Table 5F.

Table 5F. Domain Analysis of NOV5a			
Pfam Domain	NOV5a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 6.

The NOV6 clone was analyzed, and the nucleotide and encoded polypeptide
 5 sequences are shown in Table 6A.

Table 6A. NOV6 Sequence Analysis			
	SEQ ID NO: 19	556 bp	
NOV6a, CG102975-01 DNA Sequence	TGCAATATAGGGGAAAAGCAGACCATGGTGAATCCGGGCAGCAGCTCGCAGCCGCCCGGTG ACGGCCGCCCCCTCTCCTGGAAGCGGTGCGCAGGCTGCGGGGGCAAGATTGCGGACCGCTTT CTGCTCTATGCCATGGACAGCTATTGGCACAGCCGGTGCCTCAAGTGCTCCTGCTGCCAGGCG CAGCTGGGCGACATCGGCACGTCCTGTTACACCAAAGTGGCATGATCCTTTGCAGAAATGAC TACATTAGGTTATTGGAAATAGCGGTGCTTGCAGCGCTTGCAGACAGTCGATTCTCTGCGAGT GAACTCGTCATGAGGGCGCAAGGCAATGTGTATCATCTTAAGTGTGTTTACATGCTCTACCTGC CGGAATCGCTGGTCCCGGAGATCGGTTTCACTACATCAATGGCCATTTGAATCACTTCAG AGCAATCCACTACTGCCAGACCAGAAGGTCTGCTAAAAGGTCAGAGTAATGCAGAAATGCGTGC CTTTCATCTCAGATTGTTTCATCACAGGTGGATCCCATGTGTCTTCAGTAGAC		
	ORF Start: ATG at 25		ORF Stop: TAA at 475
	SEQ ID NO: 20	150 aa	MW at 16348.6kD
NOV6a, CG102975-01 Protein Sequence	MVNPGSSSQPPPVTAGPLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKSCCQAQLGDIGTS CYTKSGMILCRNDYIRLFGNSGACSACGQSIASELVMRAQGNVYHLKCFTCSTCRNRLVPGD RFHYINGHLNSLQSNPLLPDQKVC		
	SEQ ID NO: 21	619 bp	
NOV6b, CG102975-02 DNA Sequence	TTGCAGATTCGCCCTTATTGCAATATAGGGGAAAAGCAGACCATGGTGAATCCGGGCAGCAGC TCGCAGCCGCCCCCGGTGACGGCCGGCTCCCTCTCCTGGAAGCGGTGCGCAGGCTGCGGGGC AAGATTGCGGACCGCTTCTGCTCTATGCCATGGACAGCTATTGGCACAGCCGGTGCCTCAAG TGCTCCTGCTGCCAGGCGCAGCTGGGCGACATCGGCACGTCCTGTTACACCAAAGTGGCATG ATCCTTTGCAGAAATGACTACATTAGGTTATTGGAAATAGCGGTGCTTGCAGCGCTTGCAGG CAGTCGATTCTGCGAGTGAACCTCGTCATGAGGGCGCAAGGCAATGTGTATCATCTTAAGTGT TTTACATGCTCTACCTGCCGGAATCGCCTGGTCCCGGAGATCGGTTTCACTACATCAATGGC AGTTTATTTGTGAACATGATAGACCTACAGCTCTCATCAATGGCCATTTGAATCACTTCAG AGCAATCCACTACTGCCAGACCAGAAGGTCTGCTAAAAGGTCAGAGTAATGCAGAAATGCGTGC CTTTCATCTCAGATTGTTTCATCACAGGTGGATCCCATGTGTCTTCAGTAGAC		
	ORF Start: ATG at 43		ORF Stop: TAA at 538
	SEQ ID NO: 22	165 aa	MW at 17993.4kD
NOV6b, CG102975-02 Protein Sequence	MVNPGSSSQPPPVTAGPLSWKRCAGCGGKIADRFLLYAMDSYWHSRCLKSCCQAQLGDIGTS CYTKSGMILCRNDYIRLFGNSGACSACGQSIASELVMRAQGNVYHLKCFTCSTCRNRLVPGD RFHYINGSLFCEHDRPTALINGHLNSLQSNPLLPDQKVC		

Sequence comparison of the above protein sequences yields the following sequence
 relationships shown in Table 6B.

Table 6B. Comparison of NOV6a against NOV6b.		
Protein Sequence	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV6b	1..150 1..165	149/165 (90%) 149/165 (90%)

Further analysis of the NOV6a protein yielded the following properties shown in Table 6C.

Table 6C. Protein Sequence Properties NOV6a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV6a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 6D.

Table 6D. Geneseq Results for NOV6a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW61544	Human rhombotin-like protein - Homo sapiens, 165 aa. [WO9829546-A1, 09-JUL-1998]	1..150 1..165	148/165 (89%) 148/165 (89%)	5e-86
AAO08461	Human polypeptide SEQ ID NO 22353 - Homo sapiens, 112 aa. [WO200164835-A2, 07-SEP-2001]	54..150 1..112	95/112 (84%) 95/112 (84%)	7e-50
AAW78226	Fragment of human secreted protein encoded by gene 1 - Homo sapiens, 266 aa. [WO9856804-A1, 17-DEC-1998]	58..150 115..222	93/108 (86%) 93/108 (86%)	3e-49
AAM41406	Human polypeptide SEQ ID NO 6337 - Homo sapiens, 149 aa. [WO200153312-A1, 26-JUL-2001]	21..132 15..126	64/112 (57%) 81/112 (72%)	3e-37
AAM39620	Human polypeptide SEQ ID NO 2765 - Homo sapiens, 145 aa. [WO200153312-A1, 26-JUL-2001]	21..132 11..122	64/112 (57%) 81/112 (72%)	3e-37

In a BLAST search of public sequence databases, the NOV6a protein was found to have homology to the proteins shown in the BLASTP data in Table 6E.

Table 6E. Public BLASTP Results for NOV6a

Protein Accession Number	Protein/Organism/Length	NOV6a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O00158	LIM domain transcription factor LMO4 (LIM-only protein 4) (LMO-4) (Breast tumor autoantigen) - Homo sapiens (Human), and, 165 aa.	1..150 1..165	149/165 (90%) 149/165 (90%)	3e-86
Q8QG63	LIM only 4 protein - Brachydanio rerio (Zebrafish) (Zebra danio), 167 aa.	1..150 1..167	117/167 (70%) 131/167 (78%)	4e-65
Q924W9	Putative LMO1 homologue - Mus musculus (Mouse), 156 aa.	21..136 22..148	68/127 (53%) 85/127 (66%)	1e-37
Q99MB5	Neuronal specific transcription factor DAT1 - Rattus norvegicus (Rat), 155 aa.	21..132 21..132	64/112 (57%) 81/112 (72%)	4e-37
P25800	Rhombotin-1 (Cysteine rich protein TTG-1) (T-cell translocation protein 1) (LIM-only protein 1) - Homo sapiens (Human), 156 aa.	21..132 22..133	64/112 (57%) 81/112 (72%)	4e-37

5

PFam analysis predicts that the NOV6a protein contains the domains shown in the Table 6F.

Table 6F. Domain Analysis of NOV6a

Pfam Domain	NOV6a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LIM	23..82	24/62 (39%) 50/62 (81%)	2.8e-18
LIM	87..139	18/61 (30%) 42/61 (69%)	4.4e-12

10

Example 7.

The NOV7 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 7A.

Table 7A. NOV7 Sequence Analysis

	SEQ ID NO: 23	2532 bp	
NOV7a, CG103764-01 DNA Sequence	CGGCGCGCTCGCTCCCCGACCCGGACTCCCCATGTATGACGACTCCTACGTGCCCGGTT TGAGGACTCGGAGGCGGGTTCAGCCGACTCCTACACGAGCCGCCATCTCTGGACTCAGACGT CTCCCTGGAGGAGGACCGGAGAGTGCCCGCGTGAAGTAGAGAGCCAGGCTCAGCAGCAGCT CGAAAGGGCCAAGCACAAACCTGTGGCATTTCGGGTGAGGACCAATGTCAGCTACTGTGGCGT ACTGGATGAGGAGTGCCAGTCCAGGGCTCTGGAGTCAACTTTGAGGCCAAAGATTTCTGCA CATTAAAGAGAAGTACAGCAATGACTGGTGGATCGGGCGGCTAGTGAAAGAGGGCGGGACAT CGCCTTCATCCCCAGCCCCAGCGCTGGAGAGCATCCGGCTCAAACAGGAGCAGAAGGCCAG GAGATCTGGGAACCTTCCAGCCTGAGTGACATTGGCAACCGACGCTCCCTCCGCCATCTCT AGCCAAGCAGAAGCAAAGCAGGCGGAACATGTTCCCCCATATGACGTGGTGCCTCCATGCG GCCTGTGGTGTGGTGGGACCTCTCTGAAAGTTATGAGGTACAGACATGATGCAGAAGGC TCTCTTCGACTTCTCAAACACAGATTGATGGCAGGATCTCCATCACCCGAGTCACAGCCGA CCTCTCCCTGGCAAAGCGATCTGTGCTCAACAATCCGGGCAAGAGGACCATCATTGAGCGCTC CTCTGCCCGCTCCAGATTGCGGAAGTGCAGAGTGAGATCGAGCGCATATTTGAGTGGCCAA ATCCCTGCAGCTAGTAGTGTGGACGCTGACACCATCAACACCCAGCAGACTGGCCAAAGAC CTCGTGGCCCCCATCATCGTCTTTGTCAAAGTGTCTCACCAGGTAAGTCTCCAGCGTCTCAT TCGCTCCCGGGGAAGTACAGATGAAGCACCTGACCGTACAGATGATGCGATATGATAAGCT GGTTCAGTGCCACCGAGTCATTGATGTGATTCTGGATGAGAACCAGCTGGAGGATGCCTG TGAGCACCTGGCTGAGTACCTGGAGGTTACTGGCGGGCCACGACACCCAGCCCTGGCCCC CGGACTTCTGGTCTCCAGTGCCATCCCGGACTTCAGAACCCAGCAGCTGCTGGGGAGCG TGGCGAGGAGCACTCCCGCTTGAGCGGGACAGCTTGATGCCCTCTGATGAGGCCAGCGAGAG CTCCGCCAAGCCTGGACAGGATCTTCACAGCGTAGCTCCCGCCACCTGGAGGAGGACTATGC AGATGCCCTACCGAGCCTGTACCAGCCTCACCGCCAACACACCTCGGGGCTGCCTAGTGCTAA CGGGCATGACCCCCAAGACCGGCTTCTAGCCAGGACTCAGAGCACAACCACAGTGACCGGAA CTGGCAGCGCAACCGGCTTGGCCCAAGGATAGCTACTGACAGCTCCTGCTGCCCTACCCCTG GCAGGCACAGGCGCAGCTGGCTGGGGGGCCCACTCCAGGCAGGCTGGCGTTAGACTGGCATCA GGCTGGCACTAGGCTCAGCCCCAAAACCCCTGCCAGCCCACTCAGGGTGCCTGTGG TCCCAAGGTTCTGGGAGAAACAGGGGACCCCTCACCTCCTGGGCACTGACCCCTACTAGGCT CCCATTCCAGTACTAGCTGTGTCTTGCACCCCTGGCACCTTCTCTCTCCACACAGGA AGCTGCCCTCCTGGGCACTGCCCTCAGGCCAGGATCCCTTAGCAGGCTCCTTCCACACAGAC TCAGGGAAGGATGCCCCATTAAAGTGACAAAAGGGTGGGGTGTGGGCACCATGGCATGAGGA AGAAACAAGGTCCCTGAGCAGGCACAAGTCTGACAGTCAAGGGACTGCTTTGGCATCCAGGG CCTCCAGTCACCTCACTGCCATACATTAGAAATGAGACAATCAAAGCCCCCCCCAGGGTGGCA CACCCATCCGTTTGTGGGGTGTGGCAGCCACATCCAAGACTGGAGCAGGCTGGCCACGC TCGGGCCAGAGAGAGCTCACAGCTGAAGCTCTTGAGGGGAAGGGCTCTCCTCACCTGCCAGG AAGCTTCTTAACATGTGACAGGACCCAGGGACAGGAGCATGGTGAAGCCAAGTGGCAGATGGG AGCCAACCTGGATGGGGGTTTGGGGAAGGAGGGCATGTGTAGCAGAGAACTTAGGGGGGCTC CTTGCCCTTCTCATTCTTTTGGCCCTGCATCCTGTCTTCTGTTCTTGTCCCTCATACATCTT TGGAGAACCAGGCTCCAGACTTTGTTCCCTGACTCATAGCTGCGCTTGTAGGTTAGGGTTA GATGGGGAGAGACAGGGCACAGAGGACCTGTCTCCCGGCTACTCTTGCCCTTATGGCTCTAGT GTGTGACCTACAGAGCATGTCCACAAGCCCTGCCTCACCTCACTGTATCACTAATAAACA TCATGCACAGTC		
	ORF Start: at 2		ORF Stop: TGA at 1487
	SEQ ID NO: 24	495 aa	MW at 55582.3kD
NOV7a, CG103764-01 Protein Sequence	GAARSPDPSPMYDDSYVPFGEDSEAGSADSYTSRPSLSDSVLEEDRESARREVESQAQQQL ERAKHKPVFAVRNTNVSYCVLDEECPVQSGVNFKAQDFLHKEKYSNDWWIGRLVKEGGDI AFIPSPQRLESIRLKQEQKARRSGNPSSLSDIGNRRSPPPSLAKQKQAEHVPYDVVPSMR PVVLVGPLKGYEVTDMQKALFDLKHFRFDGRISITRVADLSLAKRSVLNPNPKRTIERS SARSSIAEVQSEIERIFELAKSLQLVVLDAITINHPAQLAKTSLAPIIVFVKVSSPKVLQRLI RSRGKSQMKHLTVQMAYDKLVQCPPEFDFVILDENQLEDAEHLAEVLEVYRATHHPAGGP GLLGPPSAIPGLQNLQLLGERGEEHSPLEDRSLMPSDEASESSRQAWTGSSQRRSRHLEEDYA DAYQDLYQPHRQHTSGLPSANGHPQDRLLAQDSEHNHSDRNWQRNRPWPKDSY		
	SEQ ID NO: 25	2532 bp	
NOV7b, CG103764-01 DNA Sequence	CGGCGCGCTCGCTCCCCGACCCGGACTCCCCATGTATGACGACTCCTACGTGCCCGGTT TGAGGACTCGGAGGCGGGTTCAGCCGACTCCTACACGAGCCGCCATCTCTGGACTCAGACGT CTCCCTGGAGGAGGACCGGAGAGTGCCCGCGTGAAGTAGAGAGCCAGGCTCAGCAGCAGCT CGAAAGGGCCAAGCACAAACCTGTGGCATTTCGGGTGAGGACCAATGTCAGCTACTGTGGCGT ACTGGATGAGGAGTGCCAGTCCAGGGCTCTGGAGTCAACTTTGAGGCCAAAGATTTCTGCA CATTAAAGAGAAGTACAGCAATGACTGGTGGATCGGGCGGCTAGTGAAAGAGGGCGGGACAT CGCCTTCATCCCCAGCCCCAGCGCTGGAGAGCATCCGGCTCAAACAGGAGCAGAAGGCCAG GAGATCTGGGAACCTTCCAGCCTGAGTGACATTGGCAACCGACGCTCCCTCCGCCATCTCT AGCCAAGCAGAAGCAAAGCAGGCGGAACATGTTCCCCCATATGACGTGGTGCCTCCATGCG		

	<p>GCCTGTGGTGTGGTGGGACCCTCTCTGAAAGGTTATGAGGTACACAGACATGATGCAGAAGGC TCTCTTCGACTTCCTCAAACACAGATTGATGGCAGGATCTCCATCACCCGAGTCACAGCCGA CCTCTCCCTGGCAAAGCGATCTGTGCTCAACAATCCGGGCAAGAGGACCATCATTGAGCGCTC CTCTGCCCGCTCCAGCATTGCGGAAGTGCAGAGTGAGATCGAGCGCATATTTGAGCTGGCCAA ATCCCTGCAGCTAGTAGTGTGGACGCTGACACCATCAACCACCCAGCAGAGCTGGCCAAGAC CTCGTGGCCCCCATCATCGTCTTTGTCAAAGTGTCTCACCAGGTTACTCCAGCGTCTCAT TCGCTCCCGGGGAAGTCACAGATGAAGCACCTGACCGTACAGATGATGGCATATGATAAGCT GGTTCAGTGGCCACCGGAGTCATTTGATGTGATTCTGGATGAGAACCAGCTGGAGGATGCCTG TGAGCACTGGCTGAGTACCTGGAGGTTTACTGGCGGGCCACGCACCACCCAGCCCTGGCCC CGGACTTCTGGGTCCTCCAGTGCCATCCCCGGACTTCAGAACCAGCAGCTGCTGGGGGAGCG TGGCGAGGAGCACTCCCCCTTGAGCGGGACAGCTTGATGCCCTCTGATGAGGCCAGCGAGAG CTCCCGCCAAGCCTGGACAGGATCTTACAGCGTAGCTCCCGCCACCTGGAGGAGGACTATGC AGATGCCTACCAGGACCTGTACCAGCTCACCGCCAACACACCTCGGGGCTGCCATGTGCTAA CGGGCATGACCCCCAAGACCGGCTTCTAGCCAGGACTCAGAGCACAACCACAGTGACCGGAA CTGGCAGCGCAACCGGCTTGGCCCAAGGATAGCTACTGACAGCTCTGCTGCCCTACCTG GCAGGCACAGGCGCAGCTGGCTGGGGGGCCCACTCCAGGCAGGTTGGCGTTAGACTGGCATCA GGCTGGCACTAGGCTCAGCCCCAAAACCCCTGCCAGCCAGCTTCAGGGCTGCCTGTGG TCCCAAGGTTCTGGGAGAAACAGGGGACCCCTCACCTCTGGGCAGTGACCCCTACTAGGCT CCCATTCAGGTAAGTGTGTGTTCTGCACCCCTGGCACCTTCTCTCTCCACACAGGA AGCTGCCCCACTGGGCAGTGCCCTCAGGCCAGGATCCCTTAGCAGGTTCTTCCACCAGAC TCAGGGAAGGGATGCCCCATTAAAGTGACAAAAGGGTGGGGTGTGGGCACCATGGCATGAGGA AGAAACAAGGTCCTGAGCAGGCACAAGTCTGACAGTCAAGGAGTCTTTGGCATCCAGGG CCTCCAGTCACCTCACTGCCATACATTAGAATGAGACAATCAAAGCCCCCCCCAGGGTGGA CACCCATCCGTTTGTGGGGTGTGGCAGCCACATCCAAGACTGGAGCAGCAGGCTGGCCACGC TCGGGCCAGAGAGAGCTCAGAGCTGAAGCTCTGGAGGGAAGGGCTCTCTCACCCTGCCAGG AAGCTTCTTAACATGTGACAGGACCAGGGACCAGGAGCATGGTGAAGCAAGTGGCAGATGGG AGCCAACCTGGATGGGGGTTGGGGAAGGAGGGCATGTGTAGCAGAGAACTTAGGGGGGCTC CTTGCTTTCTCATTCTTTTGCCTGCATCCTGTCTTTCTGTTCTTGTCCCTCATACATCTT TGGAGAACCGGGTCCAGACTTTGTTCCTGACTCATAGCTGCCGCTTGTAGGTAGGGTTA GATGGGGAGAGACAGGGCACAGAGGACCTGTCTCCCGGCTACTCTTGCTTATGGCTCTAGT GTGTGACCTACAGGCATGTCCACAAGCCCTGCCTCACCTCACTGTCATCAATAAACA TCATGCACAGTC</p>		
	ORF Start: at 2	ORF Stop: TGA at 1487	
	SEQ ID NO: 26	495 aa	MW at 55582.3kD
NOV7b, CG103764-01 Protein Sequence	<p>GAARSPDPSPMYDDSYVPGFEDSEAGSADSYTSRPSLSDSVSLEEDRESARREVESQAQQQL ERAKHKPVAFVRITNSYCGVLDEECVPQSGSVNFEAKDFLHIKEKYSNDWIGRLVKEGGDI AFIPSPQRLESIRLKQEQKARRSGNPSSLSIDIGNRRSPPPSLAKQKQKQAEHVPPYDVVPSMR PVVLVGPSPKGYEVTDMQKALFDLKHFRFDGRISI TRVTADLSLAKRSVLNPNPKRTIIERS SARSSIAEVQSEIERIFELAKSLQVLVDADTINHPAQLAKTSLAPIIVFVKVSSPKVLQRLI RSRGKSQMKHLTVQMMAYDKLVQCPPEFVILDENQLEDACEHLAEYLVWRATHHPAPGP GLLGPPSAIPGLQNQQLGERGEEHSPLEDSLMPSEASESSRQAWTGSSQRSSRHLEEDYA DAYQDLYQPHRQHTSGLPSANGHPQDRLLAQDSEHNHSDRNWQRNRPWPKDSY</p>		
	SEQ ID NO: 27	1822 bp	
NOV7c, 212779035 DNA Sequence	<p>TTCGTAACAACCTCCGCCCATTTGACGCAAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATA TAAGCAGAGCTCTCTGGCTAACTAGAGAACCCACTGCTTACTGGCTTATCGAAATTAATACGA CTCACTATAGGGAGACCAAGCTGGCTAGCGTTTAAACTTAAGCTTGGTACCGAGCTCGGATC CACCATGTATGACGACTCTACGTGCGCGGTTTGGAGGACTCGGAGGCGGGTTCAGCCGACTC CTACACCAGCGCCCATCTCTGAGCTCAGACGCTCCCTGGAGGAGACCGGGAGAGTGCCCG GCGTGAAGTAGAGAGCCAGGCTCAGCAGCAGCTCGAAAGGGCCAAGCACAACCTGTGGCATT TGCGGTGAGGACCAATGTGAGCTACTGTGGCGTACTGGATGAGGAGTGCCCACTCCAGGGCTC TGGAGTCAACTTTGAGGCCAAAGATTTTCTGCACATTAAGAGAAGTACAGCAATGACTGGTG GATCGGGCGGCTAGTGAAGAGGGCGGGGACATCGCCTTCATCCCGAGGACCGCCCTGGA GAGCATCCGGCTCAAACAGGAGCAGAAGGCCAGGAGATCTGGGAACCCCTTCAGCCTGAGTGA CATTGGCAACCGACGCTCCCTCCGCCATCTCTAGCCAAGCAGAAGCAAAGCAGGCGGAACA TGTTCCCCCATATGACGTGGTGCCCTCATGCGGCTGTGGTGTGGTGGGACCCTCTCTGAA AGGTTATGAGGTACAGACATGATGAGAAGGCTCTCTTCGACTTCTCAAACACAGATTTGA TGGCAGGATCTCCATCACCCGAGTACAGCCGACCTCTCCCTGGCAAGAGCATCTGTGCTCAA CAATCCGGGCAAGAGGACCATATTGAGCGCTCCTCTGCCCGCTCCAGCATTGGCGAAGTGCA GAGTGAGATCGAGCGCATATTTGAGCTGGCCAAATCCCTCGAGCTAGTAGTGTGGACGCTGA CACCATCAACCACCCAGCACAGCTGGCCAAGACCTCGCTGGCCCCCATCATCGTCTTTGTCAA AGTGTCTTCAACAAAGTACTCCAGCGTCTATTGCTCCCGGGGAAGTACAGATGAAGCA CCTGACCGTACAGATGATGGCATATGATAAGCTGGTTCAGTGCCCAACCGGAGTCATTTGATGT GATTCTGGATGAGAACCAGCTGGAGGATGCCTGTGAGCACCTGGCTGAGTACCTGGAGGTTTA CTGGCGGGCCACGCACCAACCCAGCCCTGGCCCCGGACTTCTGGGTCTCCAGTGCCATCCC</p>		

	CGGACTTCAGAACAGCAGCTGCTGGGGGAGCGTGGCGAGGAGCACTCCCCCTTGAGCGGGA CAGCTTGATGCCCTCTGATGAGGCCAGCGAGAGCTCCCGCCAAGCCTGGACAGGATCTTCACA GCGTAGCTCCCGCCACCTGGAGGAGGACTATGCAGATGCCTACCAGGACCTGTACCAGCCTCA CCGCCAACACACCTCGGGGCTGCCTAGTGCTAACGGGCATGACCCCCAAGACCGGCTTCTAGC CCAGGACTCAGAACCAACACAGTGACCGGAACCTGGCAGCGCAACCGGCTTGCGCCAAGGA TAGCTACTGAGCGGCGCTCGAGTCTAGAGGGCCCGTTAAACCCGCTGATCAGCCTCGACTG TGCCTTCTAGTTGCCAGCCTCTGTTGTTTGGCCCTCCCCGTCCTTCTTGACCTGGAAG GTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAATTGCATCGCATTGCTGAG		
	ORF Start: at 137		ORF Stop: TGA at 1646
	SEQ ID NO: 28	503 aa	MW at 56533.6kD
NOV7c, 212779035 Protein Sequence	GDPSWLAFKLKLTGELGSTMYYDSYVPGFEDSEAGSADSYTSRPSLSDSVSLEEDRESARREV ESQAQQQLERAKHKPVAFVVRTNVSYCGVLDEECVPVQSGVNFCAKDFLHIKEKYSNDWWIGR LVKEGGDIAFIPSPQRLESIRLKQEQKARRSGNPSSLSIDIGNRRSPPPSLAKQKQKQAEHVPP YDVVPSMRPVVLVGPVSLKGYEVTDMMQKALFDLKHFRDGRISITRVTDLSLAKRSVLNPNP KRTIERSARSIAEVQSEIERIFELAKSLQLVVLADTINHPAQLAKTSLAPIIVFVKVSS PKVLQRLIRSRGKSMKHLTVQMMAYDKLVQCPPESFDVILDENQLEDACEHLAEYLEVYWR THHPAPGPGLLGPPSAIPGLQNLQGLGERGEEHSPLESLMPSDEASESSRQAWTGSSQRSS RHLEEDYADAYQDLVQPHRQHTSGLPSANGHDPQDRLLAQDSEHNHSDRNWQRNRPWPKDSY		
	SEQ ID NO: 29	1449 bp	
NOV7d, CG103764-02 DNA Sequence	ATGTATGACGACTCCTACCTGCCCGGGTTTGGAGACTCGGAGGCGGGTTCAGCCGACTCCTAC ACCAGCCGCCCATCTCTGGACTCAGACGCTCTGGAGGAGACCGGAGAGTGCCCGCGGTGAA GTAGAGAGCCAGGCTCAGCAGCAGCTCGAAAGGGCCAAGCACAACTGAGGCATTGCGGTG AGGACCAATGTACGCTACTGTGGCTACTGGATGAGGAGTGCCAGTCCAGGCTCTGGAGTC AACTTTGAGGCCAAAGATTTCCTCCACATTAAAGAGAAGTACAGCAATGACTGGTGGATCGGG CGGCTAGTGAAAGAGGGCGGGGACATCGCCTTCATCCCCAGCCCCAGCGCTGGAGAGCATC CGGCTCAACAGGAACAGAAAGGAGATCCGGGAACCTTCCAGCCTGAGTGACATTGGC AACCGACGTTCCCTCTCCATCTCTAGCCAAGCAGAAGCAAAGCAGGCGGAACATGTTCC CCGTATGACGTGGTGCCCTCAATGCGGCCGCTGGTGTCTGGTGGGACCTCTCTGAAAGGTAT GAGGTACAGACATGATGCAGAAGCTCTCTCGACTTCTCAAACACAGATTGATGGCAGG ATCTCCATCACCCGAGTCACAGCCGACCTCTCCCTGGCAAAGCGATCTGTGCTCAACAATCCG GGCAAGAGGACCATATTGAGCGCTCTCTGCCCCGCTCCAGCATTGCGGAAGTGACAGTGAG ATCGAGCCCATATTTGAGCTGGCCAAATCCCTGCAGTAGTAGTGTTGGACGTGACACCATC AACCACCCAGCACAGCTGGCCAAGACCTCGCTGGCCCCCATCATCGTCTTTGTCAAAGTGTCC TCACCAAGGTACTCCAGCTCTCATTGCTCCCGGGGAAGTCACAGATGAAGCACCTGACC GTACAGATGGCATATGATAAGCTGGTTCAGTGCCCAACCGAGTCATTGATGTGATTCTGGAT GAGAACCAGCTGGAGGATGCCTGTGAGCTCTGGCTGAGTACCTGGAGGTTTACTGGCGGGCC ACGCACCAACCCAGCCCCGCGCCGAGTCTGGGTCTCCAGTGCCATCCCCGAGCTTCAG AACCAGCAGCTGCTGGGGGAGCGTGGCGAGGAGCACTCCCCCTTGAGCGGGACAGCTTGATG CCCTCTGATGAGGCGAGCGAGCTCCCGCCAAGCCTGGACAGGATCTTCACAGCGTAGCTCC CGCCACCTGGAGGAGGACTACGCAGATGCCTACCAGGACCTGTACCAGCCTACCGCCAACAC ACCTCGGGGCTGCTAGTGCTAACGGGCATGACCCCAAGACCGGCTTCTAGCCAGGACTCA GAACACAACCACAGTGACCGGAAGTGGCAGCGCAACCGCCCTTGGCCCAAGGATAGCTACTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1447
	SEQ ID NO: 30	482 aa	MW at 54347.1kD
NOV7d, CG103764-02 Protein Sequence	MYDDSYLPGFEDSEAGSADSYTSRPSLSDSVLEEDRESARREVESQAQQQLERAKHKPEAFV RTNVSYCGVLDEECVPVQSGVNFCAKDFLHIKEKYSNDWWIGRLVKEGGDIAFIPSPQRLES RLKQEQKARRSGNPSSLSIDIGNRRSPPPSLAKQKQKQAEHVPPYDVVPSMRPVVLVGPVSLKGY EVTDMMQKALFDLKHFRDGRISITRVTDLSLAKRSVLNPNPKRTIERSARSIAEVQSE IERIFELAKSLQLVVLADTINHPAQLAKTSLAPIIVFVKVSSPKVLQRLIRSRGKSMKHLTV QMMAYDKLVQCPPESFDVILDENQLEDACEHLAEYLEVYWRATHHPAPGPGLLGPPSAIPGLQ NQQLLGERGEEHSPLESLMPSDEASESSRQAWTGSSQRSSRHLEEDYADAYQDLVQPHRQHT SGLPSANGHDPQDRLLAQDSEHNHSDRNWQRNRPWPKDSY		

	SEQ ID NO: 31	1822 bp	
NOV7e, CG103764-03 DNA Sequence	TTCGTAACTCCGCCCATTTAGCGCAAATGGGCGGTAGGCGGTACGGTGGGAGGTCTATA TAAGCAGAGCTCTCTGGCTAACTAGAGAACCCTGCTTACTGGCTTATCGAAATTAATACGA CTCACTATAGGAGACCAAGCTGGCTAGCGTTAACTTAAGCTGGTACCGAGCTCGGATC CACCATGTATGACGACTCCTACGTGCCCGGGTTTGGAGACTCGGAGGCGGGTTCAGCCGACTC CTACACCAGCGGCCATCTCTGGACTCAGACGCTCTCCCTGGAGGAGGACCGGAGAGTGCCCG		

	GCGTGAAGTAGAGAGCCAGGCTCAGCAGCAGCTCGAAAGGGCCAAGCACAAACCTGTGGCATT TGCGGTGAGGACCAATGTCAGTACTGTGGCGTACTGGATGAGGAGTGGCCAGTCCAGGGCTC TGGAGTCAACTTTGAGGCCAAAGATTTCTGCACATTAAAGAGAAGTACAGCAATGACTGGTG GATCGGGCGGCTAGTGAAGAGGGCGGGGACATCGCCTTCATCCCCAGCCCCAGCGCTTGA GAGCATCCGGCTCAAACAGGAGCAGAAGGCCAGGAGATCTGGGAACCTTCCAGCCTGAGTGA CATTGGCAACCGACGCTCCCTCCGCCATCTCTAGCCAAGCAGAAGCAAAGCAGGCGGAACA TGTCCCCCATATGACGTGGTGCCCTCCATGCGGCCCTGTGGTGTGGTGGGACCTCTCTGAA AGGTTATGAGGTACAGACATGATGCAGAAGGCTCTCTTCGACTTCCTCAAACAGATTGTA TGGCAGGATCTCCATCACCGAGTCACAGCCGACCTCTCCCTGGCAAAGCGATCTGTGCTCAA CAATCCGGGCAAGAGGACCATCATTGAGCGCTCCTCTGCCCGCTCCAGCATTGCGGAAGTGCA GAGTGAGATCGAGCGCATATTTAGCTGGCCAAATCCCTGCAGCTAGTAGTGTGGACGCTGA CACCATCAACCACCCAGCAGCTGGCCAAGACCTCGCTGGCCCCCATCATCGTCTTTGTCAA AGTGTCTCACCAAAGGTAATCCAGCGTCTCATTGCTCCCGGGGAAGTCACAGATGAAGCA CCTGACCGTACAGATGATGGCATATGATAAGCTGGTTCAGTGCCCAACCGGAGTCATTGTGATG GATTCTGGATGAGAACCAGCTGGAGGATGCCTGTGAGCACCTGGCTGAGTACCTGGAGGTTTA CTGGCGGGCCACGCACACCCAGCCCTGGCCCCGGAATCTTGGGTCTCCAGTGCCATCCC CGGACTTCAGAACCAGCAGCTGCTGGGGGAGCGTGGCGAGGAGCACTCCCCCTTGAGCGGGA CAGCTTGATGCCCTCTGATGAGGCCAGCGAGAGCTCCCGCAAGCCTGGACGATCTTACACA GCGTAGCTCCCGCCACCTGGAGGAGGACTATGCAGATGCCTACCAGGACCTGTACAGCCTCA CCGCCAACACACCTCGGGGTGCTAGTGCTAACGGGCATGACCCCAAGACCGGCTTCTAGC CCAGGACTCAGAACACAACACAGTGACCGGAAGTGGCAGCGCAACCGGCCTTGGCCCAAGGA TAGCTACTGAGCGGCGCTCGAGTCTAGAGGGCCCGTTTAAACCCGCTGATCAGCCTCGACTG TGCCTTCTAGTTGCCAGCCATCTGTTGTTTGGCCCTCCCGCTGCCTTCCCTGACCCTGGAAG GTGCCACTCCCACTGTCCTTCTCTAATAAAATGAGGAAATTCATCGCATTGTCTGAG		
	ORF Start: ATG at 194		ORF Stop: TGA at 1646
	SEQ ID NO: 32	484 aa	MW at 54531.3kD
NOV7e, CG103764-03 Protein Sequence	MYDDSYVPGFEDSEAGSADSYTSRPSLSDVSLIEDRESARREVESQAQQQLERAKHKPVFAFA VRTNVSYCGVLDEECVPQSGVNFCAKDFLHIKEKYSNDWWIGRLVKEGGDIAFIPSPQRLES IRLKQEQKARRSGNPSSLDIGNRRSPPPSLAKQKQKQAEHVPPYDVVPSMRPVVLVGPPLKG YEVTDMMQKALFDLKHFRDGRISITRVTDLSLAKRSVLNPNPKRTI IERSARSSIAEVQS EIERIFELAKSLQLVVLADTINHQAQLAKTSLAPI I VFVKVSSPKVLQRLIRSRGKSQMKHL TVQMMAYDKLVQCPPEFDFVILDENQLEDACEHLAEYLEVYWRATHHPAPGPGLLGPPSAIPG LQNQQLLGERGEEHSPLERDSLMPSEASESSRQAWTGSSQRSSRHLEEDYADAYQDLYQPHR QHTSGLPSANGHDPQDRLLAQDSEHNHSDRNWQNRNPWPKDSY		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 7B.

Table 7B. Comparison of NOV7a against NOV7b through NOV7e.		
Protein Sequence	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV7b	1..495 1..495	450/495 (90%) 450/495 (90%)
NOV7c	10..495 18..503	455/486 (93%) 455/486 (93%)
NOV7d	12..495 1..482	448/484 (92%) 450/484 (92%)
NOV7e	12..495 1..484	454/484 (93%) 454/484 (93%)

5

Further analysis of the NOV7a protein yielded the following properties shown in Table 7C.

Table 7C. Protein Sequence Properties NOV7a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.1000 probability located in plasma membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV7a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several

5 homologous proteins shown in Table 7D.

Table 7D. Geneseq Results for NOV7a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW37880	Human calcium channel b3 subunit - Homo sapiens, 484 aa. [WO9811131-A2, 19-MAR-1998]	12..495 1..484	484/484 (100%) 484/484 (100%)	0.0
AAR39564	Human neuronal VDCC beta-subunit encoded by clone HBB2 - Homo sapiens, 484 aa. [DE4222126-A, 19-AUG-1993]	12..495 1..484	482/484 (99%) 484/484 (99%)	0.0
AAB10578	Human calcium channel beta3-1 subunit protein - Homo sapiens, 483 aa. [US6096514-A, 01-AUG-2000]	21..495 3..483	473/481 (98%) 473/481 (98%)	0.0
AAR76214	Human calcium channel subunit beta 3-1 - Homo sapiens, 483 aa. [WO9504822-A, 16-FEB-1995]	21..495 3..483	473/481 (98%) 473/481 (98%)	0.0
AAB10585	Human calcium channel beta-4 subunit protein #2 - Homo sapiens, 520 aa. [US6096514-A, 01-AUG-2000]	27..470 50..520	330/474 (69%) 367/474 (76%)	e-172

In a BLAST search of public sequence databases, the NOV7a protein was found to have homology to the proteins shown in the BLASTP data in Table 7E.

10

Table 7E. Public BLASTP Results for NOV7a				
Protein Accession Number	Protein/Organism/Length	NOV7a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

P54284	Dihydropyridine-sensitive L-type, calcium channel beta-3 subunit (CAB3) (Voltage-dependent calcium channel beta-3 subunit) - Homo sapiens (Human), 484 aa.	12..495 1..484	484/484 (100%) 484/484 (100%)	0.0
P54287	Dihydropyridine-sensitive L-type, calcium channel beta-3 subunit (CAB3) (Voltage-dependent calcium channel beta-3 subunit) - Rattus norvegicus (Rat), 484 aa.	12..495 1..484	480/484 (99%) 484/484 (99%)	0.0
S62185	calcium channel beta3 chain a - mouse, 484 aa.	12..495 1..484	478/484 (98%) 482/484 (98%)	0.0
AAM22963	Calcium channel beta 3 subunit - Mus musculus (Mouse), 484 aa.	12..495 1..484	479/484 (98%) 481/484 (98%)	0.0
Q9MZL3	Dihydropyridine-sensitive L-type, calcium channel beta-3 subunit (CAB3) (Voltage-dependent calcium channel beta-3 subunit) - Bos taurus (Bovine), 484 aa.	12..495 1..484	477/484 (98%) 481/484 (98%)	0.0

PFam analysis predicts that the NOV7a protein contains the domains shown in the Table 7F.

Table 7F. Domain Analysis of NOV7a			
Pfam Domain	NOV7a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Ca_channel_B	184..401	183/226 (81%) 217/226 (96%)	9.7e-182

5

Example 8.

The NOV8 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 8A.

Table 8A. NOV8 Sequence Analysis			
	SEQ ID NO: 33	2037 bp	
NOV8a, CG104944-01 DNA Sequence	CTACTTGGTCTCCTGCTTTCGCGACATGGCCTTCAATTTTGGGGCTCCCTCGGGCACCTCCGG TACCGCTGCAGCCACCGCGGCCCGGGCTGGGTTTGGAGGATTGGGACAACATCTACAACCTG CAGGTTCTGCATTGAGCTTTTCTGCCCCAACTAACACAGGCACTACTGGACTCTTTGGTGGTA CTCAGAACAAAGGTTTGGATTGTTGGTACTGTTTGGCACAACAACGGGAAGTAGTACTGGTT TAGGTACTGGTTTGGGAAGTGGACTGGGATTGGAGGATTAAATACACAGCAGCAGCAGCAAA CTAGCAGTAGGTTATAGTTGCATGCCAGTAATAAAGATGAAGATGGGCTAGTGGTTTATGTT TTCAACAAAAAGAAACAGAGATCGAAGCCAACAACAACAGTTGGTAGAATCATTGCATAAA GTTTGGGAGGAAACAGACCCTTACTGTAAATGTAGAGGGCACTAAACATTGCCAGATGAT		

	CAGACAGAAGTTGTTATTGTTGTTGAGCGTTCGCCAAATGGTACTTCAAGAAGAGTTCCA GCTACAACGCTATATGCCCAATTTTGAACAAGCCAATATAAAAAACACAATTGCAGCAACTTGGT GTAACCCCTTTCTATGACTAGAACAGAACTTTCTCCTGCACAGATCAGACAGCTTTTACAGAAT CCTCCTGCTGGTGTGATCCTATTATCTGGGAACAGGCCAAGGTAGATAACCCCTGATTCTGAA AAGTTAATTCCTGTACCAATGGTGGGTTTTAAGGAACCTTCTCCGAAGACTGAAGGTTCAAGAT CAGATGACTAAGCAGCATCAAACCAGATTAGATATCATATCTGAAGATATTAGTGAGCTACAA AAGAATCAAACACATCTGTAGCCAAAATTGCACAATACAAGAGGAAACTCATGGATCTTTCC CATAGAACTTTACAGGTCCTAATCAAACAGGAAATTCAAAGGAAGAGTGGTTATGCCATTGAG GCTGATGAAGAGCAGTTGCGAGTTCAGCTGGATACGATTGAGGGTGAACATAATGCACCTACT CAGTTCAAGGGCCGACTAAATGAATTGATGTCTCAAATCAGGATGCAGAACTATTGAGGAGCA GTCAGATCTGAAGAAAGGTATTACATAGATGCAGATCTGTTACGAGAAATCAAGCAGCATTG AAACAACAACAGGAAGGCCTTAGCCATTGATTAGCATCATTAAAGACGATCTAGAAGATATA AAGCTGGTTCGAACATGGATTGAATGAAACCATCCACATCAGAGGTGGTGTCTTTAGTTGACAG TTCACAACTTGTGTAAGGTTGTGAAATGCATCTTCTTACTGCATCAGACCTTCTCTTAAGA ATGAAACCGACCATGGAGGGAAAAAGAAAAACAATTCTTCTTGGATTGGTTTTTTGAGAAG TTTACTGACAAATTACTGTTTCATCAAATCTGAAATAGTCACCTCAGAGCTCTTCAAAGAAAA CTTTGAAAGATTTATATCTAAAGCTGTATTTACTTTAAAGAAAGTGCATAATTACCAAAT GTATGTACTATTGTACATTTTACAACAGCATTCTTCTTAAACATAATCTGTGTTTAAATGATTA TTGTCCATTGAGCCTGTACTCTGCTTCCATACCAAGTAAATATGAAATAATCTTACTTTGCAC ATAACAGAAGAACTATAATTACTTGGCTGTTGGAGATTGTACTTGAGTATAAATGTACACC AGTTTTTGTATTGTGAACTCATCTGTGGGAGGAGTAAAGAAATCCAAAGCATTTAATGTT TTGTTTTTGTCTATAAAGATATGAAATGTATTTTATATTATTTACTTATTGGAATTTA CAGAGCACACCTAAGCAATTAGGATATAACAAAACCTACTTAACCATTTTGAACCATTTTTGT TTTTTAAGCCTTTTTATTCTAAAAAGATGAAAACCTATAAATAAATCTTAATTTGTAATTA CTTTTAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 337		ORF Stop: TGA at 1318
	SEQ ID NO: 34	327 aa	MW at 37377.4kD
NOV8a, CG104944-01 Protein Sequence	MPSNKDEDGLVVLVFNKKETEIRSQQQLVESLHKVLGGNQLTVNVEGKTLPPDQTEVVIY VVERSPNGTSRRVPATTLYAHEFEQANIKTQLQLGVTLMSRTELSPAQIRQLLNPPAGVDP IIWEQAKVDNPDSEKLIPVPMVGFKEILLRRLKVQDQMTKQHQRDLIISEDISELQKNQTTSV AKIAQYKRKLMDLSHRTLQVLIKQEIQRKSGYAIQADEEQLRVQLDTIQGELNAPTQFKRLN ELMSQIRMQNHFGAVRSEERYIIDADLLREIKQHLKQQQEGLSHLISIIKDDLEDIKLVEHGL NETHIRGGVFS		

Further analysis of the NOV8a protein yielded the following properties shown in Table 8B.

Table 8B. Protein Sequence Properties NOV8a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1590 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV8a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 8C.

Table 8C. Geneseq Results for NOV8a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV8a	Identities/ Similarities for	Expect Value

		Match Residues	the Matched Region	
AAB42518	Human ORFX ORF2282 polypeptide sequence SEQ ID NO:4564 - Homo sapiens, 354 aa. [WO200058473-A2, 05-OCT-2000]	1..327 28..354	327/327 (100%) 327/327 (100%)	0.0
AAB60098	Human transport protein TPPT-18 - Homo sapiens, 507 aa. [WO200078953-A2, 28-DEC-2000]	1..327 181..507	326/327 (99%) 327/327 (99%)	0.0
AAB93036	Human protein sequence SEQ ID NO:11814 - Homo sapiens, 484 aa. [EP1074617-A2, 07-FEB-2001]	1..302 181..482	297/302 (98%) 301/302 (99%)	e-166
ABB64020	Drosophila melanogaster polypeptide SEQ ID NO 18852 - Drosophila melanogaster, 610 aa. [WO200171042-A2, 27-SEP-2001]	1..317 288..606	125/320 (39%) 193/320 (60%)	4e-62
AAB42664	Human ORFX ORF2428 polypeptide sequence SEQ ID NO:4856 - Homo sapiens, 237 aa. [WO200058473-A2, 05-OCT-2000]	1..56 182..237	56/56 (100%) 56/56 (100%)	8e-25

In a BLAST search of public sequence databases, the NOV8a protein was found to have homology to the proteins shown in the BLASTP data in Table 8D.

Table 8D. Public BLASTP Results for NOV8a				
Protein Accession Number	Protein/Organism/Length	NOV8a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P70582	Nucleoporin p54 - Rattus norvegicus (Rat), 510 aa.	1..327 184..510	315/327 (96%) 325/327 (99%)	e-179
Q96EA7	Unknown (protein for MGC:13407) - Homo sapiens (Human), 493 aa.	1..313 181..493	312/313 (99%) 313/313 (99%)	e-175
Q9P011	Nucleoporin p54 protein - Homo sapiens (Human), 505 aa.	1..327 180..505	311/327 (95%) 315/327 (96%)	e-173
Q9NVL5	CDNA FLJ10655 fis, clone NT2RP2005933, weakly similar to nucleoporin NUP57 - Homo sapiens (Human), 484 aa.	1..302 181..482	297/302 (98%) 301/302 (99%)	e-166
Q9V6B9	CG8831 protein (LD24841P) - Drosophila melanogaster (Fruit fly), 610 aa.	1..317 288..606	125/320 (39%) 193/320 (60%)	1e-61

PFam analysis predicts that the NOV8a protein contains the domains shown in the Table 8E.

Table 8E. Domain Analysis of NOV8a

Pfam Domain	NOV8a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 9.

The NOV9 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 9A.

Table 9A. NOV9 Sequence Analysis

	SEQ ID NO: 35	1471 bp	
NOV9a, CG106550-01 DNA Sequence	<p>TTGAACATGGACGAAGGAATTCTCTATTGTCAGAGAGACAGTTACTGGAACATAGAGATTTT ATAGGACTTGGACTATTCTCTTTGTATGTGTAACCCAAAAGGAGCATGAAACGAGACGAA ACCAAGGATACCTACAAATTACCGCACAGATTAATAGAAAAGAAAAGACCGAATTAAT GAATGCATTGCTCAGCTGAAAGATTTACTGCCTGAACATCTGAAATTGACAACCTCTGGGACAT CTGGAGAAAGCTGTAGTCTTGAATTAACCTTTGAAACACTTAAAGCTTTAACCGCCTTAAC GAGCAACAGCATCAGAAGATAATTGCTTTACAGAATGGGGAGCGATCTCTGAAATCGCCCAT CAGTCCGACTTGGATTGCGTTTCCACTCGGGATTTCAAACATGCGCCAAAGAAAGTCTTGCAATG CTCTCCCGGTTTGAGAGCTGGACACCCAGGGAGCGCGGTGTGTCCAGCTGATCAACCACTT CAGCCCGTGGCCACCCAGTTCTTGCCACCCCGCAGCTGTTGACTCAACAGGTCCCTCTGAGC AAAGGCACCGGCGCTCCCTCGGCCCGCGGTCGCGGCCGCCCTGCTGGAGCGCGCGGG CAGAAGCTGGAGCCCCCTCGCTACTGCGTGCCCGTCATCAGCGGACTCAGCCAGCGCCGAA CTCGCCGCGCAGACGACACGACAGCGAGCTACGGCGGCGAAGCCGAGGCCCGGCC GACCGCGAGAAAGGCAAAGGCGCGGGGGCGAGCCGCGTCACCATCAAGCAGGAGCCTCCCGG GAGGACTCGCCGCGCCCAAGAGGATGAAGCTGGATTCCCGCGCGCGCGCAGCGCGCGCGG CCGGGGGGCGCGCGCGCGCGCGCGCAGCCGCGCTTCTGGGGCCCGAACCTGCCCGCGCGG CGCTCTGAGACCCGACGCCGCCCTGCTCAGCTCGCTGTTGGCGTTTCGGCGGAGCGGAGG GGCGCCCTTCCCGCAGCCCGCGCGCGCGCGCCCTTCTGCTGCTTCTGCTTCTCTCTCC CCTTCTGCAGCTGCCGCTACGTGCAGCCCTTCTGGACAAGAGCGGCTGGAGAAGTATCTG TACCGCGCGCGGCTGCGGCCCGGTTCCCGCTGCTATACCCCGGCATCCCGCCCCGCGCGCA GCCGCGGACCGCCGCCGCCGCTGCCGCCGCCGCCCGCGTCCCTGCTGCTCTCGGT TTGTGCGCCCCCTCCGAGAAGCGCGCGCGCGCGGCAACCTTCGCGCACGAGGTGGCG CCCCCTTGGGGCGCGCACCCCCAGCACCCGCAGCGCGCACCCACTTCCCTTTCGCGGGCC CGCGAGCCGGGAACCCGAGAGCTCTGCTCAGGAAGATCCCTCGCAGCCAGGAAGGAAGCT CCTTGAATCCTTGCGTCCCGAA</p>		
	ORF Start: ATG at 7		ORF Stop: TGA at 1453
	SEQ ID NO: 36	482 aa	MW at 50496.7kD
NOV9a, CG106550-01 Protein Sequence	<p>MDEGIPLQERQLEHRDFIGLDYSSLYMCKPKRSMKRDDTKDYLPHRLIEKKRRDRINEC IAQLKDLLPEHLKLTTLGHLEKAVVLELTLKHLKALTALTEQQHQKIALONGERSLKSPIQS DLDAFHSGFQTCAKEVLQYLSRFESWTPREPRCVQLINHLHAVATFLPTPOLLTQOVPLSKG TGAPSAAGSAAAPKLERAGQKLEPLAYCVPVIQRTQPSAELAAENDTDDTSYGYGGEAEARPDR EKGKGAGASRVITIKQEPGEDSPAPKRMKLDLRGGSGGGPGGGAAGAAAALLPDPAAAAAL LRPDAALLSSLVAFGGGGGAPFPQAAAAAPFLPFCFLSPSAAAAAYVQPLDKSGLEKLYLP AAAAAPFPLLYPGIPAPAAAAAATAAAAAAFAFLSSSVLSPPEKAGAAAATLLPHEVAPL GAPHPQHPHGRTHLPFAGPREPGNPSSAQEDPSQPGKEAP</p>		
	SEQ ID NO: 37	628 bp	
NOV9b,	TTGAACATGGACGAAGGAATTCTCTATTGTCAGAGAGACAGTTACCGGAACATAGAGATTTT		

CG106550-02 DNA Sequence	ATAGGACTGGACTATTCTCTTTGTATATGTGTAAACCCAAAAGGAGCATGAAACGAGACGAC ACCAAGGATACCTACAAATTACCGCACAGATTAATAGAAAAGAAAAGAGACCGAATTAAT GAATGCATTGCTCAGCTGAAAGATTACTGCCTGAACATCTGAAATTGACAACCTCTGGGACAT CTGGAGAAAGCTGTAGTCTTGAATTAACCTTTGAAACACTTAAAGCTTTAACCGCCTTAACC GAGCAACAGCATCAGAAGATAATTGCTTTACAGAATGGGGAGCGATCTCTGAAATCGCCCAT CAGTCCGACTTGGATGCGTTCCACTCGGGATTTCAAACATGCGCCAAAGAAGTCTTGAATAC CTCTCCCGGTTTGAGGGCTGGACACCCAGGGAGCCGCGGTGTGTCCAGCTGATCAACCACTTG CACGCCGTGGCCACCCAGTTCTGCGCTTCGCGGGGCCCCGCGAGCCGGGGAACCCGGAGAGC TCTGCTCAAGAAGATCCCTCGCAGCCAGGAAAGGAAGCTCCCTGAATCCTTGGTCCCGAA		
	ORF Start: at 1		ORF Stop: TGA at 610
	SEQ ID NO: 38	203 aa	MW at 23298.4kD
NOV9b, CG106550-02 Protein Sequence	LNMDEGI PHLQERQLPEHRDFIGLDYSSLYMCKPKRSMKRDDTKDITYKLPHRLIEKKRRDRIN ECIAQLKDLLPEHLKLTTLGHLEKAVVLELTCLKHLKALTALTEQQHQKIIALQNGERSLKSPI QSDLDAPHSGFQTCAKEVLQYLSRFEGWTPREPRCVQLINHLHAVATQFLPFAGPREPGNPES SAQEDPSQPGKEAP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 9B.

Table 9B. Comparison of NOV9a against NOV9b.		
Protein Sequence	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV9b	1..193	163/201 (81%)
	3..203	165/201 (81%)

Further analysis of the NOV9a protein yielded the following properties shown in Table 9C.

Table 9C. Protein Sequence Properties NOV9a	
PSort analysis:	0.7000 probability located in plasma membrane; 0.3000 probability located in nucleus; 0.2000 probability located in endoplasmic reticulum (membrane); 0.1000 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV9a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 9D.

Table 9D. Geneseq Results for NOV9a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value

AAB70692	Human DEC2a protein sequence SEQ ID NO:2 - Homo sapiens, 482 aa. [WO200114551-A1, 01-MAR-2001]	1..482 1..482	482/482 (100%) 482/482 (100%)	0.0
AAB70693	Human DEC2b protein sequence SEQ ID NO:12 - Homo sapiens, 484 aa. [WO200114551-A1, 01-MAR-2001]	1..482 1..484	482/484 (99%) 482/484 (99%)	0.0
AAB70694	Mouse DEC2a protein sequence SEQ ID NO:14 - Mus musculus, 410 aa. [WO200114551-A1, 01-MAR-2001]	1..482 1..410	350/484 (72%) 371/484 (76%)	e-178
AAU16188	Human novel secreted protein, Seq ID 1141 - Homo sapiens, 165 aa. [WO200155322-A2, 02-AUG-2001]	37..201 1..165	165/165 (100%) 165/165 (100%)	1e-90
AAU16603	Human novel secreted protein, Seq ID 1556 - Homo sapiens, 150 aa. [WO200155322-A2, 02-AUG-2001]	37..186 1..150	148/150 (98%) 148/150 (98%)	9e-81

In a BLAST search of public sequence databases, the NOV9a protein was found to have homology to the proteins shown in the BLASTP data in Table 9E.

Table 9E. Public BLASTP Results for NOV9a				
Protein Accession Number	Protein/Organism/Length	NOV9a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9C0J9	Class B basic helix-loop-helix protein 3 (bHLHB3) (Differentially expressed in chondrocytes protein 2) (hDEC2) (Enhancer-of-split and hairy-related protein 1) (SHARP-1) - Homo sapiens (Human), 482 aa.	1..482 1..482	482/482 (100%) 482/482 (100%)	0.0
Q8TAT1	Basic helix-loop-helix domain containing, class B, 3 - Homo sapiens (Human), 482 aa.	1..482 1..482	481/482 (99%) 481/482 (99%)	0.0
Q99PV5	Class B basic helix-loop-helix protein 3 (bHLHB3) (Differentially expressed in chondrocytes protein 2) (mDEC2) - Mus musculus (Mouse), 410 aa.	1..482 1..410	350/484 (72%) 371/484 (76%)	e-177
O35779	Class B basic helix-loop-helix protein 3 (bHLHB3) (Enhancer-of-split and hairy-related protein 1) (SHARP-1) - Rattus norvegicus (Rat), 410 aa.	1..482 1..410	348/484 (71%) 371/484 (75%)	e-176

O35185	Class B basic helix-loop-helix protein 2 (bHLHB2) (Stimulated with retinoic acid 13) (E47 interaction protein 1) (eipl) - Mus musculus (Mouse), 411 aa.	14..401 20..364	179/405 (44%) 224/405 (55%)	4e-66
--------	---	--------------------	--------------------------------	-------

PFam analysis predicts that the NOV9a protein contains the domains shown in the Table 9F.

Table 9F. Domain Analysis of NOV9a			
Pfam Domain	NOV9a Match Region	Identities/ Similarities for the Matched Region	Expect Value
HLH	48..100	14/57 (25%) 39/57 (68%)	0.0055

5

Example 10.

The NOV10 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 10A.

Table 10A. NOV10 Sequence Analysis			
	SEQ ID NO: 39	1344 bp	
NOV10a, CG106842-01 DNA Sequence	TCGTCCGCAAAGCCTGAGTCCTGTCCTTTCTCTCTCCCGGACAGCATGAGCTTCACCACTCG CTCCACCTTCTCCACCACTACCGGTCCCTGGGCTCTGTCCAGGCGCCAGCTACGGCGCCCG GCCGGTCAGCAGCGCGGCCAGCGTCTATGCAGGCGCTGGGGCCCTGGCCACCGGGATAGCCGG GGGTCTGGCAGGAATGGGAGGCATCCAGAACGAGAAGGAGACCATGCAAAGCCTGAACGACCG CCTGGCCTCTTACTTGGACAGAGTGAGGAGCCTGGAGACCGAGAACCGGAGGCTGGAGAGCAA AATCCGGGAGCACTTGGAGAAGAAGGGACCCAGGTGAGAGACTGGAGCCATTACTTCAAGAT CATCGAGGACCTGAGGCTCAGATCTTCGAAATACTGTGGACAATGCCCCGATCGTTCTGCA GATTGACAATGCCCGTCTTGCTGCTGATGACTTTAGAGTCAAGTATGAGACAGAGCTGGCCAT GCGCCAGTCTGTGGAGAACGACATCCATGGGCTCCGCAAGGTCAATTGATGACACCAATATC ACGACTGCAGCTGGAGACAGAGATCGAGGCTCTCAAGGAGGAGCTGCTCTTCATGAAGAAGAA CCACGAAGAGGAAGTAAAGGCCCTACAAGCCAGATTGCCAGCTCTGGGTTGACCGTGGAGGT AGATGCCCCCAATCTCAGGACCTCGCCAAGATCATGGCAGACATCCGGGCCCAATATGACGA GCTGGCTCGGAAGAACCAGAGGAGCTAGACAAGTACTGGTCTCAGCAGATTGAGGAGAGCAC CACAGTGGTCACCACACAGTCTGCTGAGGTTGGAGCTGCTGAGACGACGCTCAGAGCTGAG ACGTACAGTCCAGTCTTGGAGATCGACCTGGACTCCATGAGAAATCTGAAGGCCAGCTTGA GAACAGCCTGAGGGAGGTGGAGGCCCGCTACGCCCTACAGATGGAGCAGCTCAACGGGATCCT GCTGCACCTTGAGTCAAGAGCTGGCAGACACCGGGCAGAGGGACAGCGCCAGGCCAGGAGTA TGAGGCCCTGCTGAACATCAAGGTCAAGCTGGAGGCTGAGATCGCCACCTACCGCCCGCTGCT GGAAGATGGCGAGGACTTAACTTGGTGATGCCTTGGACAGCAGCAACTCCATGCAAAACCT CCAAAAGACCACCACCGCCGGATAGTGGATGGCAAAGTGGTGTCTGAGACCAATGACACCAA AGTTCTGAGGCATTAAGCCAGCAGAAGCAGGTACCCTTGGGGAGCAGGAGGCCAATAAAAA GTTGAGAGTTCATTGGATGTC		
	ORF Start: ATG at 47		ORF Stop: TAA at 1274
	SEQ ID NO: 40	409 aa	MW at 46045.1kD
NOV10a, CG106842-01 Protein Sequence	MSFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASVYAGAGGLATGIAGGLAGMGGIQNEKETM QSLNDRLASYLDRVRSLETENRRLESKIREHLEKKGQVRDWSHYFKIIEDLRAQIFANTVDN ARIVLQIDNARLAADDFRVKYETELAMRQSVENDIHGLRKVIDDTNITRLQLETEIEALKEEL LFMKNHEEEVKGQLAQIASSGLTVEVDAPKSQLAKIMADIRAQYDELARKNREELDKYWSQ QIEESTTVVTTQSAEVGAAETTLTELRRTVQSLEIDLDSMRNLKASLENSLREVEARYALQME		

	QLNGILLHLESELAQTRAEGQRQAQYEALLNIKVKLEAEIATYRRLLEDGEDFNLGDALDSS NSMQTIQKTTTRRIVDGKVVSETNDTKVLRH
	SEQ ID NO: 41 1412 bp
NOV10b, CG106842-02 DNA Sequence	CGGGGTCGTCGCGCAAAGCCTGAGTCCTGTCTCTCTCTCCCCGGACAGCATGAGCTTCACC ACTCGCTCCACCTTCTCCACCAACTACCGGTCCCTGGGCTCTGTCCAGGCGCCAGCTACGGC GCCCCGCGGTCAGCAGCGCGGCCAGCGTCTATGCAGGCGCTGGGGGCTCTGGTTCCTCGGATC TCCGTGTCCCGCTCCACCAGCTTCAAGGGCGGCATGGGGTCCGGGGGCTGGCCACCGGGATA GCCGGGGGTCTGGCAGGAATGGGAGGCATCCAGAACGAGAAGGAGACCATGCAAAGCCTGAAC GACCGCCTGGCCTCTTACCTGGACAGAGTGAGGAGCCTGGAGACCGAGAACCGGAGGCTGGAG AGCAAAATCCGGGAGCACTTGGAGAAGAAGGACCCAGGTCAGAGACTGGAGCCATTACTTC AAGATCATCGAGGACCTGAGGGCTCAGATCTTCGCAAACTACTGTGGACAATGCCCGCATCGTT CTGCAGATTGACAATGCCGCTCTTGTCTGTGATGACTTAGAGTCAAGTATGAGACAGAGCTG GCCATGCGCCAGTCTGTGGAGAACGACATCCATGGGCTCCGCAAGGTCATTGATGACACCAAT ATCACACGACTGCAGCTGGAGACAGAGATCGAGGCTCTCAAGGAGGAGCTGCTCTTCATGAAG AAGAACCACGAAGAGGAAGTAAAAGGCCTACAAGCCAGATTGCCAGCTCTGGGTGACCGTG GAGGTAGATGCCCCAAATCTCAGGACCTCGCAAGATCATGGCAGACATCCGGGCCCAATAT GACGAGCTGGCTCGGAAGAACCGAGAGGAGCTAGACAAGTACTGGTCTCAGCAGATTGAGGAG AGCACCACAGTGGTCACCACACAGTCTGCTGAGGTTGGAGCTGCTGAGACGACGCTCACAGAG CTGAGACGTACAGTCCAGTCTTGGAGATCGACCTGGACTCCATGAGAAATCTGAAGGCCAGC TTGGAGAACAGCCTGAGGGAGGTGGAGGCCGCTACGCCCTACAGATGGAGCAGCTCAACGGG ATCCTGCTGCACCTTGAGTCAGAGCTGGCACAGACCCGGGCAGAGGGACAGCGCCAGGCCAG GAGTATGAGGCCCTGCTGAACATCAAGGTCAAGCTGGAGGCTGAGATCGCCACCTACCGCCGC CTGCTGGAAGATGGCGAGGACTTTAATCTTGGTGATGCCTTGGACAGCAGCAACTCCATGCAA ACCATCCAAAAGACCACCACCCGCCGATAGTGGATGGCAAAGTGGTGTCTGAGACCAATGAC ACCAAAGTCTGAGGCATTAAAGCCAGCAGAAGCAGGTACCCTTTGGGGAGCAGGAGGCCAAT AAAAAGTTCAGAGTTCATTGGATGTC
	ORF Start: ATG at 52 ORF Stop: TAA at 1342
	SEQ ID NO: 42 430 aa MW at 48057.2kD
NOV10b, CG106842-02 Protein Sequence	MSFTTRSTFSTNYRSLGSVQAPSYGARPVSSAASVYAGAGGSGSRI SVSRSTSFRGGMGSGGL ATGIAGGLAGMGGIQNEKETMQSLNDRLASYLDRVRSLETENRRLESKIREHLEKKGPQVRDW SHYFKIIEDLRAQIFANTVDNARIVLQIDNARLAADDFRVKYETELAMRQSVENDIHGLRKVI DDTNITRLQLETEIEALKEELLFMKNHEEEVKGLQAQIASSGLTVEVDAPKSDLAKIMADI RAQYDELARKNREELDKYWSQQIEESTTVVTTQSAEVGAAETTLTELRTVQSLEIDLDSMRN LKASLENSLREVEARYALQMEQLNGILLHLESELAQTRAEGQRQAQYEALLNIKVKLEAEIA TYRRLLEDGEDFNLGDALDSSNSMQTIQKTTTRRIVDGKVVSETNDTKVLRH

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 10B.

Table 10B. Comparison of NOV10a against NOV10b.		
Protein Sequence	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV10b	1..409 1..430	381/430 (88%) 381/430 (88%)

5

Further analysis of the NOV10a protein yielded the following properties shown in Table 10C.

Table 10C. Protein Sequence Properties NOV10a	
PSort analysis:	0.8477 probability located in mitochondrial intermembrane space; 0.7065 probability located in mitochondrial matrix space; 0.3907 probability located in mitochondrial inner membrane; 0.3907 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV10a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 10D.

5

Table 10D. Geneseq Results for NOV10a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB90795	Human shear stress-response protein SEQ ID NO: 90 - Homo sapiens, 430 aa. [WO200125427-A1, 12-APR-2001]	1..409 1..430	409/430 (95%) 409/430 (95%)	0.0
AAG74328	Human colon cancer antigen protein SEQ ID NO:5092 - Homo sapiens, 452 aa. [WO200122920-A2, 05-APR-2001]	1..409 23..452	409/430 (95%) 409/430 (95%)	0.0
ABG16550	Novel human diagnostic protein #16541 - Homo sapiens, 447 aa. [WO200175067-A2, 11-OCT-2001]	1..409 18..447	409/430 (95%) 409/430 (95%)	0.0
ABG15224	Novel human diagnostic protein #15215 - Homo sapiens, 456 aa. [WO200175067-A2, 11-OCT-2001]	1..409 24..456	388/433 (89%) 397/433 (91%)	0.0
ABG08564	Novel human diagnostic protein #8555 - Homo sapiens, 449 aa. [WO200175067-A2, 11-OCT-2001]	1..409 16..449	392/434 (90%) 394/434 (90%)	0.0

In a BLAST search of public sequence databases, the NOV10a protein was found to have homology to the proteins shown in the BLASTP data in Table 10E.

Table 10E. Public BLASTP Results for NOV10a				
Protein Accession Number	Protein/Organism/Length	NOV10a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

S05481	keratin 18, type I, cytoskeletal - human, 430 aa.	1..409 1..430	409/430 (95%) 409/430 (95%)	0.0
P05783	Keratin, type I cytoskeletal 18 (Cytokeratin 18) (K18) (CK 18) - Homo sapiens (Human), 429 aa.	2..409 1..429	408/429 (95%) 408/429 (95%)	0.0
Q96GD2	Similar to keratin 18 - Homo sapiens (Human), 375 aa (fragment).	38..409 4..375	371/372 (99%) 372/372 (99%)	0.0
P05784	Keratin, type I cytoskeletal 18 (Cytokeratin 18) (Cytokeratin endo B) (Keratin D) - Mus musculus (Mouse), 422 aa.	2..409 1..422	359/422 (85%) 385/422 (91%)	0.0
I59463	keratin, type I, cytoskeletal - mouse, 423 aa.	1..409 1..423	358/423 (84%) 385/423 (90%)	0.0

Pfam analysis predicts that the NOV10a protein contains the domains shown in the Table 10F.

Table 10F. Domain Analysis of NOV10a			
Pfam Domain	NOV10a Match Region	Identities/ Similarities for the Matched Region	Expect Value
filament	58..370	158/360 (44%) 283/360 (79%)	4.4e-157

5

Example 11.

The NOV11 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 11A.

Table 11A. NOV11 Sequence Analysis			
	SEQ ID NO: 43	1197 bp	
NOV11a, CG107095-01 DNA Sequence	CCACTGGGACATATGTGGTGTTCCTTCTAGCTCCTGTCTCCTCCTCATGCCTTTGTTGGGTA TGGGCATGTTAGGGGGAAGGTCATTGCTGTCTAGAGGGGCACTGACTTTCTAATGGTGTACCC AAGGTGAATGTTGGAGACACAGTCGCGATGCTGCCCAAGTCCCGCGAGCCCTAACTATCCAG GAGATCGCTGCGCTGGCCAGGTCCTCCCTGCATGGTATTTCCAGGTGGTGAAGGACCACGTG ACCAAGCCTACCGCCATGGCCAGGGCCGAGTGGCTCACCTCATTGAGTGAAGGGCTGGAGC AAGCCGAGTGACTCACCTGCTGCCCTGGAATCAGCCTTTTCTCCTATTAGACCTCAGCGAG GGCGAACAAGAGGCTCGCTTTGCAGCAGGAGTGGCTGAGCAGTTTGCCATCGCGGAAGCCAAG CTCCGAGCATGGTCTTCGGTGGATGGCGAGGACTCCACTGATGACTCCTATGATGAGGACTTT GCTGGGGGAATGGACACAGACATGGCTGGGCAGCTGCCCTGGGGCCGACCTCCAGGACCTG TTCACCGCCACCGGTTCTCCCGCCTGTGCGCCAGGGCTCCGTGGAGCCTGAGAGCGACTGC TCACAGACCATGTCCCCAGACACCTGTGCTCTAGTCTGTGCAGCCTGGAGGATGGGTTGTTG GGCTCCCGGGCCCGGCTGGCCTCCAGCTGCTGGGCGATGAGCTGCTTCTCGCCAACTGCCC CCCAGCCGGGAAAGTGCTTCCGCAGCTGGGCCCACTGGAGGCCAGGACTCACTTACAAC TCGCCCTCACAGATCCTGCTTTCCCCCGGAGGAGGAGCCAGCCCCCTGCAAGGACTGC CAGCCACTCTGCCACCACTAACGGGCAGCTGGGAACGGCAGCGGCAAGCCTCTGACCTGGCC TCTTCTGGGGTGGTGTCTTAGATGAGGATGAGGCAGAGCCAGAGGAACAGTGACCCACATCA		

	TGCCTGGACAGTGACCCACATCATGCCTGGACAGTGACCCACATCATGCCTGGACAGTGACCC ACATCATCCTGGACAGTGACCCACATCATGCCTGGACAGTGACCCACATCATGCCTGGACAGT GACCCACATCATGCCTGGACAGTGACCCACATCATCCTGGACAGTGACCCACATCATGCCTGG		
	ORF Start: ATG at 115		ORF Stop: TGA at 997
	SEQ ID NO: 44	294 aa	MW at 31502.5kD
NOV11a, CG107095-01 Protein Sequence	MVLPKVNVGDTVAMLPKSRRLTIQEIAALARSSLHGISQVVKDHVTKPTAMAQGRVAHLIEW KGWSKPSDSPAALESFAFSSYSDLSEGEQEARFAAGVAEQFAIAEAKLRAWSSVDGEDSTDDSY DEDFAGGMDTDMAGQLPLGPHLQDLFTGHRFSRPVRQGSVEPESDCSQTMSPDTLCSSLCSLE DGLLGSPARLASQLLGDELLAKLPPSRESAFRSLGLEAQSLYNSPLTESCLSPAEEEPAP CKDCQPLCPPLTGSWERQRQASDLASSGVVSLDEDEAEPEEQ		

Further analysis of the NOV11a protein yielded the following properties shown in Table 11B.

Table 11B. Protein Sequence Properties NOV11a	
PSort analysis:	0.3600 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV11a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 11C.

Table 11C. Geneseq Results for NOV11a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU81979	Human secreted protein SECP5 - Homo sapiens, 328 aa. [WO200198353-A2, 27-DEC-2001]	1..294 35..328	293/294 (99%) 294/294 (99%)	e-169
AAM79122	Human protein SEQ ID NO 1784 - Homo sapiens, 366 aa. [WO200157190-A2, 09-AUG-2001]	3..294 75..366	290/292 (99%) 292/292 (99%)	e-167
AAG81313	Human AFP protein sequence SEQ ID NO:144 - Homo sapiens, 281 aa. [WO200129221-A2, 26-APR-2001]	14..294 1..281	279/281 (99%) 280/281 (99%)	e-161
AAM80106	Human protein SEQ ID NO 3752 - Homo sapiens, 382 aa. [WO200157190-A2, 09-AUG-2001]	3..294 91..382	267/292 (91%) 272/292 (92%)	e-153

AAB66099	Protein of the invention #11 - Unidentified, 335 aa. [WO200078961-A1, 28-DEC-2000]	37..294 78..335	257/258 (99%) 258/258 (99%)	e-149
----------	--	--------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV11a protein was found to have homology to the proteins shown in the BLASTP data in Table 11D.

Table 11D. Public BLASTP Results for NOV11a				
Protein Accession Number	Protein/Organism/Length	NOV11a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8TA84	Similar to unknown (Protein for IMAGE:3487926) - Homo sapiens (Human), 281 aa.	14..294 1..281	280/281 (99%) 281/281 (99%)	e-161
CAC38563	Sequence 143 from Patent WO0129221 - Homo sapiens (Human), 281 aa.	14..294 1..281	279/281 (99%) 280/281 (99%)	e-160
Q99KY5	Hypothetical 16.1 kDa protein - Mus musculus (Mouse), 150 aa (fragment).	143..290 2..149	130/148 (87%) 138/148 (92%)	4e-71
Q9D390	6330503C03Rik protein - Mus musculus (Mouse), 300 aa.	10..290 3..297	120/318 (37%) 159/318 (49%)	6e-38
O94871	KIAA0773 protein - Homo sapiens (Human), 300 aa.	10..290 3..297	120/309 (38%) 160/309 (50%)	2e-37

5

Pfam analysis predicts that the NOV11a protein contains the domains shown in the Table 11E.

Table 11E. Domain Analysis of NOV11a			
Pfam Domain	NOV11a Match Region	Identities/ Similarities for the Matched Region	Expect Value
lacI	20..47	7/28 (25%) 18/28 (64%)	0.68

10

Example 12.

The NOV12 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 12A.

Table 12A. NOV12 Sequence Analysis			
	SEQ ID NO: 45	538 bp	
NOV12a, CG107477-01 DNA Sequence	CATCTGAAGCCAAAGAGATAAACTTTATGCCAGATTCCCCCTATAGAGAAGATGGATGCATC CTTGTCCATGCTTGCTAATTGCGAGAAGCTTTCAGTGTCTACAACTGCATTGAAAAAATTGC CAACCTGAATGGCTTAAAAAAGCTTGAAGATATTATCTTTAGGAAGAAACAACATAAAGAAGCTT AAATGGACTGGAGGCTGTAGGGGACACATTAGAAGAACTGTGGATCTCCTACAATTTTATTGA GAAGTTGAAAGGGATCCACATAATGAAGAAATTGAAGATTCTCTACATGTCTAATAACCTGGT AAAAGACTGGGCTGAGTTTGTGAAGTTGGCAGAACTGCCATGCCTCGAAGACCTGGTGTGTTGT AGGCAATCCCTTGAAGAGAAACATTCTGCTGAGAATAACTGGATTGAAGAAGCAACCAAGAG AGTGCCCAACTGAAAAGCTGGATGGTACTCCAGTAATTAAGGGGATGAGGAAGAAGACAA CTAATGCCACGCTTTCCACATGTGTGTTAACTTA		
	ORF Start: ATG at 53		ORF Stop: TAA at 506
	SEQ ID NO: 46	151 aa	MW at 17095.6kD
NOV12a, CG107477-01 Protein Sequence	MDASLSMLANCEKLSLSTNCIEKIANLNLKLNLRILSLGRNNIKNLNGLEAVGDTLEELWISY NFIEKLKGIHIMKKLKILYMSNNLVKDWAEFVKLAELPCLEDLVFVGNPLEEKHSAENNWIEE ATKRVPKLKKLDGTPVIKGDDEEDN		
	SEQ ID NO: 47	633 bp	
NOV12b, CG107477-02 DNA Sequence	AGTAGCAACCGCCGGAATGGCGAAAGCAACAACATCAAAGAAGCCTTAGCGAGATGGGAAGA GAAAACTGGCCAGAGGCCATCTGAAGCCAAAGAGATAAACTTTATGCCAGATTCCCCCTAT AGAGAAGATGGATGCATCCTTGTCCATGCTTGCTAATTGCGAGAAGCTTTCAGTGTCTACAA CTGCATTGAAAAAATTGCCAACCTGAATGGCTTAAGAGGCAGTAGGGGACACATTAGAAGAAC TGTGGATCTCCTACAATTTTATTGAGAAGTTGAAAGGGATCCACATAATGAAGAAATTGAAGA TTCTCTACATGTCTAATAACCTGGTAAAAGACTGGGCTGAGTTTGTGAAGCTGGCAGAACTGC CATGCCTCGAAGACCTGGTGTGTTGTAGGCAATCCCTTGAAGAGAAACATTCTGCTGAGAATA ACTGGATTGAAGAAGCAACCAAGAGAGTGCCCAACTGAAAAGCTGGATGGTACTCCAGTAA TTAAAGGGGATGAGGAAGAAGACAATAATGCCACGCTTTCCACTGTGTGTTAACTTATTTAA ATGTCATAAGAACAATAGATAAAATTTATATAATTGTCTATTTTAAAAAATAAAAAAAAAA AAA		
	ORF Start: ATG at 17		ORF Stop: TGA at 275
	SEQ ID NO: 48	86 aa	MW at 9691.1kD
NOV12b, CG107477-02 Protein Sequence	MAKATTIKEALARWEEKTGQRPSEAKEIKLYAQIPPIEKMDASLSMLANCEKLSLSTNCIEKI ANLNLGRSGRGIIRRTVDLLQFY		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 12B.

Table 12B. Comparison of NOV12a against NOV12b.		
Protein Sequence	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV12b	1..34 40..73	31/34 (91%) 32/34 (93%)

5

Further analysis of the NOV12a protein yielded the following properties shown in Table 12C.

Table 12C. Protein Sequence Properties NOV12a	
PSort analysis:	0.4859 probability located in mitochondrial matrix space; 0.4500 probability located in cytoplasm; 0.1967 probability located in mitochondrial inner membrane; 0.1967 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV12a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 12D.

5

Table 12D. Geneseq Results for NOV12a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU74331	Human cytoskeleton-associated protein (CYSKP) #2 - Homo sapiens, 190 aa. [WO200185942-A2, 15-NOV-2001]	1..151 40..190	151/151 (100%) 151/151 (100%)	2e-83
AAB41987	Human ORFX ORF1751 polypeptide sequence SEQ ID NO:3502 - Homo sapiens, 196 aa. [WO200058473-A2, 05-OCT-2000]	1..151 46..196	151/151 (100%) 151/151 (100%)	2e-83
ABG12773	Novel human diagnostic protein #12764 - Homo sapiens, 144 aa. [WO200175067-A2, 11-OCT-2001]	13..151 6..144	139/139 (100%) 139/139 (100%)	2e-76
ABB59218	Drosophila melanogaster polypeptide SEQ ID NO 4446 - Drosophila melanogaster, 188 aa. [WO200171042-A2, 27-SEP-2001]	1..149 40..187	85/149 (57%) 113/149 (75%)	7e-44
AAG00733	Human secreted protein, SEQ ID NO: 4814 - Homo sapiens, 113 aa. [EP1033401-A2, 06-SEP-2000]	1..87 27..113	87/87 (100%) 87/87 (100%)	2e-43

In a BLAST search of public sequence databases, the NOV12a protein was found to have homology to the proteins shown in the BLASTP data in Table 12E.

Table 12E. Public BLASTP Results for NOV12a				
Protein Accession Number	Protein/Organism/Length	NOV12a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q9BS43	Similar to RIKEN cDNA 1700010H15 gene - Homo sapiens (Human), 151 aa.	1..151 1..151	151/151 (100%) 151/151 (100%)	5e-83
Q9DAH9	1700010H15Rik protein - Mus musculus (Mouse), 151 aa.	1..151 1..151	140/151 (92%) 146/151 (95%)	3e-77
Q8T888	Leucine-rich repeat dynein light chain - Ciona intestinalis, 190 aa.	1..151 40..190	120/151 (79%) 132/151 (86%)	2e-64
O44230	Outer arm dynein light chain 2 - Anthocidaris crassispina (Sea urchin), 199 aa.	1..151 47..197	110/151 (72%) 129/151 (84%)	1e-57
Q9V573	CG8800 protein - Drosophila melanogaster (Fruit fly), 188 aa.	1..149 40..187	85/149 (57%) 113/149 (75%)	2e-43

PFam analysis predicts that the NOV12a protein contains the domains shown in the Table 12F.

Table 12F. Domain Analysis of NOV12a			
Pfam Domain	NOV12a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 13.

The NOV13 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 13A.

Table 13A. NOV13 Sequence Analysis			
	SEQ ID NO: 49	1644 bp	
NOV13a, CG108707-01 DNA Sequence	GATATCCCGAGATTAGGTCCCCAGCTTCCAAAGAGAGGATCAGAATGTCTCAGGATAATGACA CATTGATGAGAGACATCCTGGGGCATGCGCTCGCTGCTATGAGGCTGCAGAAGCTGGAACAGC AGCGGCGGCTGTTGAAAAGAAGCAGCGACAGAAAGCGCCAGGAGCTCCTCATGGTTTCAGGCCA ATCCTGACGCTTCCCCGTGGCTTTGGCGCTCTTGTCTGCGGGAGGAGCGCCTTTTAGGTGACA GAGGCCTTGGGAACCTTTCCTCCGGAAGAAAGTGTGAGGACACATCTGCCCTCTGGCATCC ACAGTGCCCTGGGCACCGTGAGCTGTGGTGGAGACGGCAGGGGCGAGCGCGGCTCCCCGACAC CGCGGACAGAAGCAGTGTTCAGGAATCTCGGTCTCCAGTCCCCTTCTTATCCTGGCTCCCAG ACAATTCGATGCAGAATTGGAGGAAGTCTCCGTGGAGAATGGTCCGTCTCTCCCCACCTT TTAAACAGTCTCCGAGAATCCGACGCAAGGGTTGGCAAGCCCACCAACGACCTGGGACCCGTG CAGAGGTGAGAGTACTCCAGGATATGGGAGATGCACACAAGTCACCCAATATGGGACCAA ACCCTGGAATGGATGGTGACTGTGTATATGAAACTTGGCCTTCCAAAAGGAAGAAGACTTGG AAAAGAAGAGAGAGGCCTCTGAGTCTACAGGGACGAATCCTCAGCAGCACACAACGAAGAGT TGTCGAAGGCCCTGAAAGGCGAGGGTGGCACGGACAGCGACCATATGAGGCACGAAGCCTCCT TGGCAATCCGCTCCCCCTGCCCTGGGCTGGAGGAGGACATGGAAGCCTACGTGCTGCGGCAG CGCTCCCGGGCACCATGATGCAGTGCTACCTCACCCTGACAAGCACGGCGTGGACAAGGGCT TGTTCCCCCTCTACTACCTCTACCTGGAGACCTCTGACAGCCTGCAGCGCTCCCTCCTGGCTG GGCGAAAGAGAAGAAGGAGCAAACTTCTAATTACCTCATCTCCCTGGATCCTACACACCTAT CTCGGACGGGACAATTCGTGGGCAAGTCAGATCCAATGTCTTCAGCACCAAGTTTACCA		

	TCTTTGACAATGGGGTGAATCCTGACCGGGAGCATTTAACCAGGAATACTGCCCGGATCAGAC AGGAGCTGGGGGCTGTGTGTTATGAGCCCAACGTCTTAGGATACCTGGGGCCTCGGAAAATGA CTGTGATTCTCCCAGGAACCAACAGCCAGAACAGCGAATCAATGTCCAGCCACTAAATGAAC AGGAGTCGCTACTGAGTCGTTACCAACGTGGGGACAAACAAGGGTTGCTTTTGTGTCACAACA AAACCCCGTCGTGGGACAAGGAGAACGGTGTCTACACGCTCAATTTCTATGGTCGAGTCACTC GGGCTTCGGTGAAGAACTTCCAATCGTGGATCCCAAACACCGTGAGCTCCTGGAAACAGATT TAGCCGGGCCAGAGAACATCTGGTGCTCCAGTTCGGCCGAGTGGGCCAGACACATTACCA TGGACTTCTGCTTTCATTAGCCCGCTCCAGGCCTTCAGCATCTGCTTGTCCAGTTTCAATT AGAAGC		
	ORF Start: ATG at 45		ORF Stop: TAG at 1638
	SEQ ID NO: 50	531 aa	MW at 59739.7kD
NOV13a, CG108707-01 Protein Sequence	MSQDNDTLMRDILGHALAAMRLQKLEQQRRLEFEKKQRQKQELLMVQANPDASPLWRSCLRE ERLLGDRGLGNPFLRKKVSEAHLPSTGSIHSALGTVSCGGDGRGERGLPTPRTEAVFRNLGLQSP FLSWLPDNSDAELEEVSVENGVSPPPFKQSPRIIRKQWQAHQRPQTRAEGESDSQDMGDAH SPNMGPNPMDGDCVYENLAFQKEEDLEKKREASESTGTNSSAAHNEELSKALKGEGGTDSDH MRHEASLAIRSPCGLEEDMEAYVLRPALPGTMMQCYLTRDKHGVKGLFPLYLYLETSDSL QRSLLAGRKRRRSKTSNYLISLDPHTLSRDGDNFVGKVRNVFSTKFTIFDNGVNPDRHLTR NTARIRQELGAVCYEPNVLYLGPVKMTVILPGTNSQNRINVQPLNEQESLLSRYQRGDKQG LLLLHNKTPSWDKENGVTYTLNFGYGRVTRASVKNFQIVDPKHRELLETSLAGPEEHLVLQFGRV GPDFTMTDFCFPFSPLOAFSICLSSFN		

Further analysis of the NOV13a protein yielded the following properties shown in Table 13B.

Table 13B. Protein Sequence Properties NOV13a	
PSort analysis:	0.6000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV13a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 13C.

Table 13C. Geneseq Results for NOV13a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB26906	Human TULP2 protein - Homo sapiens, 520 aa. [US6114502-A, 05-SEP-2000]	1..531 1..520	515/531 (96%) 517/531 (96%)	0.0
AA W36491	Human TULP2 protein - Homo sapiens, 520 aa. [WO9738004-A1, 16-OCT-1997]	1..531 1..520	515/531 (96%) 517/531 (96%)	0.0

AAB26908	Mouse TULP4 protein - Mus sp, 506 aa. [US6114502-A, 05-SEP-2000]	45..531 1..500	267/516 (51%) 326/516 (62%)	e-126
AAW36494	Human TULP4 protein - Homo sapiens, 506 aa. [WO9738004-A1, 16-OCT-1997]	45..531 1..500	267/516 (51%) 326/516 (62%)	e-126
AAB26901	Mouse tub Form II protein - Mus sp, 505 aa. [US6114502-A, 05-SEP-2000]	5..531 7..499	200/547 (36%) 299/547 (54%)	1e-86

In a BLAST search of public sequence databases, the NOV13a protein was found to have homology to the proteins shown in the BLASTP data in Table 13D.

Table 13D. Public BLASTP Results for NOV13a				
Protein Accession Number	Protein/Organism/Length	NOV13a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8TC50	Tubby like protein 2 - Homo sapiens (Human), 520 aa.	1..531 1..520	516/531 (97%) 518/531 (97%)	0.0
O00295	Tubby related protein 2 (Tubby-like protein 2) - Homo sapiens (Human), 520 aa.	1..531 1..520	515/531 (96%) 517/531 (96%)	0.0
P46686	Tubby related protein 2 (Tubby-like protein 2) (P4-6 protein) - Mus musculus (Mouse), 564 aa (fragment).	2..531 16..558	292/559 (52%) 359/559 (63%)	e-140
S42728	phosphodiesterase (clone p4-6) - mouse, 271 aa.	262..531 7..265	175/270 (64%) 205/270 (75%)	4e-95
P50586	Tubby protein - Mus musculus (Mouse), 505 aa.	5..531 7..499	200/547 (36%) 299/547 (54%)	3e-86

5

PFam analysis predicts that the NOV13a protein contains the domains shown in the Table 13E.

Table 13E. Domain Analysis of NOV13a			
Pfam Domain	NOV13a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Tub	279..531	127/307 (41%) 238/307 (78%)	1e-212

10

Example 14.

The NOV14 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 14A.

Table 14A. NOV14 Sequence Analysis			
	SEQ ID NO: 51	3937 bp	
NOV14a, CG108791-01 DNA Sequence	TCGCTGACAGCAGCCATGGCGAGCGGCAGTGGAGACAGCGTCACCCGTCGGAGCGTGGCATC ACAGTTTTTCACTCAAGAGGAGGGGCGGGCATCGATGGCATGACCACCTCAGAGAGGGTGGT GGATCTTCTGAACGAGCGGTGCTGATCACCATGACTCAAAGATCACAGTGTCAAACAGGT CCAGGAGCTGATCATCAACAAAGACCCACACTACTGGACAACCTCCTGGATGAGATCATCGC ATTCCAAGCAGACAAGTCAATCGAAGTGCAGAAATTTGTATCGGCTTCATCGAGGAGGCATG CAAGCGAGACATCGAGTTGCTGCTGAAACTCATTGCAAACCTCAACATGCTCTTGAGGGACGA GAATGTGAACGTGGTGAAGAAGGCTATCCTCACCATGACCCAGCTCTACAAGGTGGCCCTGCA GTGGATGGTAAAGTACGGGTCAATAGCGAGCTACAGGAGGCTGCTGGGACATGGTATCTGC CATGGCGGGGACATCATCTGCTATTGGACTCTGACAATGACGGCATCCGCACCCACGCCAT CAAGTTTGTGGAGGGCTCATTGTCAACCCGTCACCCCGCATGGCTGACTCAGAGATACCCCG ACGCCAGGAGCATGATATCAGCCTGGACCGCATCCCTCGTGACCAACCTACATCCAGTACAA CGTGCTATGGGAAGAGGGCAAGGCAGCCTTGGAGCAGCTGCTTAAGTTATCGGTGCACCCCTGC CATCTCCTCATCAACCTGACCACAGCGCTGGGCTCCCTTGCCAATATCGCCCCCAGAGACC CATGTTTATGTTCTGAGGTGATCCAGGCCTATGAAACTCTGCATGCCAACCTGCCCCGACGCT GGCCAAATCGCAGGTGAGCAGTGTGCGTAAGAATCTGAAGCTGCACCTGTTGAGTGTGTGAA GCACCCGGCTTCCTTGGAGTTCCAGGCCAGATCACCACCTGCTGGTGGACCTGGGCACACC TCAGGCCGAGATCGCCCCGAACATGCCGAGCAGCAAGGACACCCGCAAGCGGCCCGCGATGA CTCGACTCCACACTCAAGAAGATGAAGCTGGAGCCCAACCTGGGGGAGGACGATGAGGACAA AGACTTGGAGCCAGGCCCCGTCGGGGACCTCGAAGGCTCAGCGCAGATCTCCGGCCAGTCAGA CACGGACATCACAGCTGAGTTCTCTCAGCCTCTGTGTGACGCTGATAATGTGGCTAATCTGGT CCTCATCAGCATGGTGATCTACCCGAGGCCATGCCAGCCTCCTTCCAGGCCATCTACACCC CGTGAGTCAGCAGGCACGGAAGCCAGATCAAGCACCTGGCTCGGCTCATGGCCACACAGAT GACAGCTGCCGACTGGGACCAGGTGTAGAGCAGACCAACAGTGCAAGGAGGAGCCCAAGGA GGAGAAGGTGGTGAAGACAGAGAGCGTCTGATCAAGCGGCGCTGTGAGCCAGGGCCAAAGC CATCTCGGTGGTGGGTTCCCTGAGCTCCATGTCCCCCTGGAGGAAGAGGCACCCGAGGCCAA GAGGAGGCCAGAGCCCATTTATCCCTGTCACTCAGCCCCGGCTGGCAGGCGCTGGTGGCGCAA GAAATTTTCCGTCTCAGCGACGTGCTGAAGCCCCCTTACCGATGCCAGGTGGAAGCCATGAA GCTGGGCGCTGTGAAGCGGATCCTGCGGGCTGAGAAGGCTGTGGCCTGCAGCGGGGACGCCA GGTCCGCATAAAGATCCTGGCCAGCCTGGTGACACAGTTCAACTCGGCGCTGAAGGCGGAGGT CCTGTCTTCATCCTGGAGATGTGCGGGCCGCTGGACCTGGCTTCGCTGGCTCTACCA GGAGTACAACGCCTACCTGGCCGAGGTGCTCGGGCTCCCTGGACAAGTATGAGGACTGCCT CATCCGCTGTGTCTGGCTGCAGGAGAAACAGACCAGAAGGATGGGATCTTACCAAGGT GTGCTGGAGGCGCCACTCATCAGAGAGTGCCTGGAGGTGGTCCGCAAGTACTGCGAGGA TGAGAGTCGCACCTATCTGGGCATGTCCACACTTCGAGACCTGATCTTCAAGCGCCGTCCTCC CCAGTTCCAGTACTGCTGATGCTCCTCGACCTCAGCTCCCATGAGAAGGACAAGGTGCGCTC CCAGGCCCTGCTGTTTCATCAAACGCATGTATGAGAAGGAGCAGCTGCGGGAGTATGTGGAGAA ATTTGCCCTCAACTACCTGCAGCTCCTGGTGACCCCCAACCCACCGTCTGTGCTGTTTGGAGC TGACAAGGACACAGAGGTGGCAGCACCTGGACGGAGGAGACAGTGAAGCAGTGTCTGTACCT CTACCTGGCCCTCCTGCCCTCAGAACCACAAGCTGATCCACGAAGTGGCGGCCGTGTACACTGA AGCCATCGCCGACATCAAGCGGACGGTGTGAGGGTCATTGAGCAGCCGATCCGAGGAATGGG CATGAATCCCGGAGCTGCTCCTGCTGGTGGAAAATTGTCCCAAGGGAGCAGAGACACTGGT CACGAGATGTCTGCACAGCTCAGACAAAAGTCCACCCCTCCCAAGAGCTGGTGAAGCGGGT CCGGGATCTTACCACAAGCGACTGCCAGAGCTCCGCTTCTCATCCCGTGCTCAATGGGCT GGAGAAGAAAGAGGTGATCCAGGCCCTGCCTAAACTCATCAAACCTCAACCCCATCGTGGTGAA GGAAGTCTTAAACCGCTGCTGGGCACCCAGCATGGTGGAGGAAACTCAGCCTTGTCCCCGCT GAACCCCTGGAGAGCTCCTGATCGATTACACAACATTGACTCCGTGAAGTGCGACATGAAATC CATCATCAAAGCCACCAACCTGTGCTTTGCGGAGCGGAACGTGTACAGCTCAGAGGTGCTGGC CGTGGTATGCAGCAGCTGATGGAGCAGAGCCCCCTGCCATGCTGCTCATGAGGACCGTCAT CCAGTCCCTGACCATGTACCCCGCCTGGGGGGCTTCGTATGAACATCTGTCCCCCCTCAT CATGAAGCAGGTGTGGAAGTACCCCAAGGTGTGGGAGGGCTTCATCAAGTGTGCCAGCGCAC AAAGCCCCAGAGCTTCCAGGTATCCTGCAGCTGCCGCCCCAGCAGCTGGGAGCCGTCTTTGA CAAGTGCCAGAGCTCCGGGAGCCCCCTGCTGGCCCATGTCCGCTCCTTACCCCCCACCAGCA AGCTCACATCCCTAACTCCATCATGACCATCTTGAGGGCCAGCGGCAAGCAGGAGCCAGAGGC CAAGGAGGCGCTGCGGGGCCCTTGGAGGAGGATGATCTGGAGCCCTGACCTTGGCCCCGGC CCCAGCACCCCGCCCCCTCAGGACCTCATCGGCTGCGACTGGCCAGGAGAAGGCCTTAA		

	GCGGCAGCTGGAGGAGGAACAGAAGCTGAAGCCGGGAGGAGTGGGAGCCCCCTCCTCTTCCTC CCCCTCTCCCTCTCCGTCCGCCCGGCCAGGCCCGCCCCCGTCTGAGGAAGCCATGGATTTCCG GGAGGAGGGGCTGAGTGCAGACCCCGGCATCTTCATCAGCATGGATGACGACTCGGGGCT GACCGAGGCCGCGCTGTGGACTCTAGTCTCGAGGGCCCCCTACCCAAGGAGACGGCAGCGGG CGGGCTGACCTTGAAGGAGGAGCGGAGCCCCCAGACCCTCGCACCTGTTGGAGAAGATGCTAT GAAGACTCCCAGCCCGCTGCCGAGGACGCCAGGGAACCCGAGGCCAAGGGGAACAGCTGACG GGGCTCGAGGGGAAAGGGGTGGGACAGGGACTCGGGGCTGGGGGACGGGGCGGGGCTTGAC CTGCGGGTGCTTTGCCTTAAAAAGAAATAAA		
	ORF Start: ATG at 17		ORF Stop: TGA at 3839
	SEQ ID NO: 52	1274 aa	MW at 141158.5kD
NOV14a, CG108791-01 Protein Sequence	MASGSGDSVTRRSVASQFFTEEGPGIDGMTTSERVVDLLNQAVLITNDSKITVLKQVQELII NKDPTLLDNFLDEIIAFQADKSIEVRKFVIGFIEEACKRDIELLLKLIANLMLLRDENVN VKKAILTMTQLYKVALQWMVKSRISEIQEACWDMVSAMAGDIILLDSNDNGIRTHAIKFVEG LIVTPSPRMADSEIPRRQEHDISLDRIPRDHPYIQYNVLWEEGKAALQLKFMVHPAIISSIN LTTALGSLANIRQRPMFMSEVIQAYETLHANLPPTLAKSQVSSVRKNLKLHLSVLKHPASL EFQAQITTTLLVDLGTQAEIARNMPSSKDTRKRPRDDSDSTLKKMKLEPNLGEDDEDKDL EPGPSGTSKASQISGQSDTDITAEFLQPLLPDNDVANLVLSMVYLPEAMPASFQAIYTPVESAG TEAQIKHLARLMATQMTAAGLPGVEQTKQCKEEPKEEKVVKTESVLIKRRLSAQGQAI SVVGLSSMSPLEEEAPQAKRRPEPIIPVTQPRLAGAGGRKKIFRLSDVLKPLTDAQVEAMKLGAVK RILRAEKAVACSGAAQVRIKILASLVTQFNSGLKAEVLSFILEDVRLRLDLAFAWLYQEY NAYLAAGASGSLDKYEDCLIRLLSGLQEKPDQKDGIFTKVVLEAPLITESALEVVRKYCE DESRTYLGMSTLRDLIFKRPSRQFQYLHVLLDSSHEKDKVRSQALLFIKMYEKEQLREY VEKFAFNLYQLLVHPNPPSVLFGADKDEVAAPWTEETVKQCLYLYLALLPQNHKLIH ELAAVYTEAIADIKRTVLRVIEQPIRGMGMNSPELLLLVENC PKGAETLVTRCLHSL TDKVPSPPELVKRVRLDLYHKRLPDVRFILPVLNGLEKKEVIQALPKLIKLNPIVVKEV FNRLGLTQHGEQNSALSPNPGELIALHNIDSVKCDMKSIIKATNLCFAERNVYTSEVLAV VMQQLMEQSPPLMLLMRTVIOQLTMYPRLGGFVMNLSRLIMKQVWKYPKVWEGFIKCCQRT KPQSFQVILQLPPQQLGAVFDKCPREPLLAHVRSFTPHQQAHPNSIMTILEASGKQEP EAKEAPAGPLEEDDLEPLTAPAPAPRPQDLIGLRLAQEKALKRQLEEQKLKPGGVGAP SSSSPSPSPSARPGPPPSEAMDFREEGPECETPGIFISMDDDSGLTEAALLDSSLEG PLPKETAAGGLTLKEERSPQTLAPVGEDAMKTPSPAEDAREPEAKGNS		

Further analysis of the NOV14a protein yielded the following properties shown in Table 14B.

Table 14B. Protein Sequence Properties NOV14a	
PSort analysis:	0.8528 probability located in nucleus; 0.5806 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.2922 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV14a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 14C.

Table 14C. Geneseq Results for NOV14a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB58861	Drosophila melanogaster polypeptide SEQ ID NO 3375 - Drosophila melanogaster, 1116 aa. [WO200171042-A2, 27-SEP-2001]	7..1103 2..1102	411/1118 (36%) 634/1118 (55%)	0.0
AAG38593	Arabidopsis thaliana protein fragment SEQ ID NO: 47634 - Arabidopsis thaliana, 1250 aa. [EP1033405-A2, 06-SEP-2000]	553..1113 612..1228	178/625 (28%) 310/625 (49%)	2e-66
AAG38592	Arabidopsis thaliana protein fragment SEQ ID NO: 47633 - Arabidopsis thaliana, 1291 aa. [EP1033405-A2, 06-SEP-2000]	553..1113 653..1269	178/625 (28%) 310/625 (49%)	2e-66
AAG38591	Arabidopsis thaliana protein fragment SEQ ID NO: 47632 - Arabidopsis thaliana, 1371 aa. [EP1033405-A2, 06-SEP-2000]	553..1113 733..1349	178/625 (28%) 310/625 (49%)	2e-66
AAB86463	Murine HCN2 protein - Mus sp, 863 aa. [WO200159153-A2, 16-AUG- 2001]	1012..1208 651..850	48/204 (23%) 75/204 (36%)	0.005

In a BLAST search of public sequence databases, the NOV14a protein was found to have homology to the proteins shown in the BLASTP data in Table 14D.

Table 14D. Public BLASTP Results for NOV14a

Protein Accession Number	Protein/Organism/Length	NOV14a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q92797	Symplekin - Homo sapiens (Human), 1142 aa.	133..1274 1..1142	1141/1142 (99%) 1141/1142 (99%)	0.0
AAH30214	Hypothetical 58.9 kDa protein - Homo sapiens (Human), 533 aa.	1..533 1..533	531/533 (99%) 531/533 (99%)	0.0
AAM49961	LD45768p - Drosophila melanogaster (Fruit fly), 1165 aa.	7..1134 2..1151	422/1166 (36%) 649/1166 (55%)	0.0
Q9VNH4	CG2097 protein - Drosophila melanogaster (Fruit fly), 1116 aa.	7..1103 2..1102	411/1118 (36%) 634/1118 (55%)	0.0
Q9D990	4632415H16Rik protein - Mus musculus (Mouse), 304 aa.	985..1274 1..304	246/307 (80%) 263/307 (85%)	e-136

PFam analysis predicts that the NOV14a protein contains the domains shown in the Table 14E.

Table 14E. Domain Analysis of NOV14a			
Pfam Domain	NOV14a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 15.

The NOV15 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 15A.

Table 15A. NOV15 Sequence Analysis			
	SEQ ID NO: 53	959 bp	
NOV15a, CG109247-01 DNA Sequence	CGGCACGAGCATGGCTACCTCAGAGCTGAGCTGCGAGGTGTCGGAGGAGAACTGTGAGCGCCG GGAGGCCCTTCTGGGCAGAATGGAAGGATCTGACACTGTCCACACGGCCCGAGGAGGGCTGCTC CCTGCATGAGGAGGACACCCAGAGACATGAGACCTACCACCAGCAGGGGCAGTGCCAGGTGCT GGTGACGCGCTCGCCCTGGCTGATGATGCGGATGGGCATCCTCGGCCGTGGGCTGCAGGAGTA CCAGCTGCCCTACCAGCGGTACTGCCGCTGCCATCTTACCCCTGCCAAGATGGGCGCCAC CAAGGAGGAGCGTGAGGACACCCCATCCAGCTTCAGGAGCTGTGGCGCTGGAGACAGCCCT GGGTGGCCAGTGTGTGGACCGCCAGGAGGTGGCTGAGATCACAAGCAGCTGCCCCCTGTGGT GCCTGTCAGCAAGCCCGGTGCACTTCGTGCGCTCCCTGTCCCGCTCCATGTCCCAGGAAGACA GAGAGGCTGAGAGGACTGTGACTTGGGCTCCGCTGTGCCCGCCCTGGGCTGGGCCCTTCTCTG GCTAGGACTGTGGAGGGGAGCTGCTGGCCATGGCTGCTTGTAGTTTGCCAGAGTTGGGGGC TAGGGGAGGGGGAGCCAGAGGCCAGGATGCCTGAGCCCCCTGAGTTCCCAAAGGGAGGGTGG CAGAGACAGTGGGCACTAAGGTGGAGAGTTGGGGGCCAGCACAGCTGAGGACCTCAGCCCC AGGAGAAGGGACAAAAGTACTGGTGAGGGCAAGAGGTGCCTGGGAGGAGTGGCCCTGATCCA GGAAAATGTGAGGGGAATCTGGAACGCTTAGGCAGAAGAAGCTGGGAGGGAGGGGAGGTGA AAAGGGCAGAGCAAGGATGGTGGGGCCCCAGCACCTCTGTTAGTGCCCAATAAATGCTC AATCATGTGCCAGA		
	ORF Start: ATG at 11		ORF Stop: TGA at 512
	SEQ ID NO: 54	167 aa	MW at 19051.5kD
NOV15a, CG109247-01 Protein Sequence	MATSELSCEVSEENCERREAFWAEWKDLTLSTRPEEGCSLHEEDTORHETYHQQQQOVVLVQR SPWLMRMGILGRGLQEYQLPYQRVLPPIFTPAKMGATKEEREDTPIQLQELLALETALGGQ CVDRQEVAEITKQLPPVVPVSKPGALRRSLSRMSQEAQRG		
	SEQ ID NO: 55	672 bp	
NOV15b, CG109247-02 DNA Sequence	CGGCACGAGCATGGCTACCTCAGAGCTGAGCTGCGAGGTGTCGGAGGAGAACTGTGAGCGCCG GGAGGCCCTTCTGGGCAGAATGGAAGGATCTGACACTGTCCACACGGCCCGAGGAGGGCTGCTC CCTGCATGAGGAGGACACCCAGAGACATGAGACCTACCACCAGCAGGGGCAGTGCCAGGTGCT GGTGACGCGCTCGCCCTGGCTGATGATGCGGATGGGCATCCTCGGCCGTGGGCTGCAGGAGTA CCAGCTGCCCTGGGCTGGGCCCTTCTGGCTAGGACTGTGGAGGGGAGCTGCTGGCCATGGCT GCTTTGTAGTTTGCCAGAGTTGGGGGCTAGGGGAGGGGGAGCCAGAGGCCAGGATGCCTGA GCCCCCTGAGTTCCCAAAGGGAGGGTGGCAGAGACAGTGGGCACTAAGGTGGAGAGTTGGGG GCCAGCACAGCTGAGGACCTCAGCCCCAGGAGAAGGGACAAAAGTACTGGTGAGGGCAAGA GGTGCTGGGAGGAGTGGCCCTGATCCAGGAAAATGTGAGGGGAATCTGGAACGCTTAGGCA GAAGAAGCTGGGAGGGAGGGGAGGTGAAAAGGCAGAGGCAAGGATGGTGGGGCCCCAGCA CCCTCTGTTAGTGCCCAATAAATGCTCAATCATGTGCCAGA		
	ORF Start: ATG at 11		ORF Stop: TAG at 344

	SEQ ID NO: 56	111 aa	MW at 12856.4kD
NOV15b, CG109247-02 Protein Sequence	MATSELSCEVSEENCERREAFWAEWKDLTLSTRPEEGCSLHEEDTORHETYHQQGQCQVLVQR SPWLMMRMGILGRGLQEYQLPWAGPFLARTVEGSCWPWLLCSLPRVGG		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 15B.

Table 15B. Comparison of NOV15a against NOV15b.		
Protein Sequence	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV15b	1..85 1..85	84/85 (98%) 85/85 (99%)

5

Further analysis of the NOV15a protein yielded the following properties shown in Table 15C.

Table 15C. Protein Sequence Properties NOV15a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV15a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 15D.

Table 15D. Geneseq Results for NOV15a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM40716	Human polypeptide SEQ ID NO 5647 - Homo sapiens, 170 aa. [WO200153312-A1, 26-JUL-2001]	1..167 4..170	167/167 (100%) 167/167 (100%)	1e-94
AAM38930	Human polypeptide SEQ ID NO 2075 - Homo sapiens, 113 aa. [WO200153312-A1, 26-JUL-2001]	1..106 1..106	106/106 (100%) 106/106 (100%)	2e-59

AAW71560	Human hepatocyte nuclear factor 1 alpha (R131Q mutant) - Homo sapiens, 630 aa. [WO9811254-A1, 19-MAR-1998]	43..121 90..164	20/79 (25%) 39/79 (49%)	1.6
AAW71562	Human hepatocyte nuclear factor 1 alpha (truncated mutant) - Homo sapiens, 415 aa. [WO9811254-A1, 19-MAR-1998]	43..121 90..164	19/79 (24%) 39/79 (49%)	4.7
AAW71561	Human hepatocyte nuclear factor 1 alpha (truncated mutant) - Homo sapiens, 314 aa. [WO9811254-A1, 19-MAR-1998]	43..121 90..164	19/79 (24%) 39/79 (49%)	4.7

In a BLAST search of public sequence databases, the NOV15a protein was found to have homology to the proteins shown in the BLASTP data in Table 15E.

Table 15E. Public BLASTP Results for NOV15a

Protein Accession Number	Protein/Organism/Length	NOV15a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O15273	Telethonin (Titin cap protein) - Homo sapiens (Human), 167 aa.	1..167 1..167	167/167 (100%) 167/167 (100%)	4e-94
Q96L27	Titin-cap (telethonin) - Homo sapiens (Human), 167 aa.	1..167 1..167	166/167 (99%) 166/167 (99%)	3e-93
O70548	Telethonin (Titin cap protein) - Mus musculus (Mouse), 167 aa.	1..167 1..167	151/167 (90%) 163/167 (97%)	4e-86
O70549	Telethonin - Mus musculus (Mouse), 167 aa.	1..167 1..167	150/167 (89%) 162/167 (96%)	1e-85
T18863	hypothetical protein C02D4.2 - Caenorhabditis elegans, 501 aa.	29..62 308..341	13/34 (38%) 21/34 (61%)	3.4

5

PFam analysis predicts that the NOV15a protein contains the domains shown in the Table 15F.

Table 15F. Domain Analysis of NOV15a

Pfam Domain	NOV15a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 16.

The NOV16 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 16A.

Table 16A. NOV16 Sequence Analysis			
	SEQ ID NO: 57	2067 bp	
NOV16a, CG110410-01 DNA Sequence	CGAGAGGAGAGCGCGAGAGCCCCAGCCGCGGGCGGGCGGGCGGTGAAGATGGCAGAGGCACCG GCTTCCCCGGCCCCGCTCTCTCCGCTCGAAGTGGAGCTGGACCCGGAGTTCGAGCCCCAGAGC CGTCCGCGATCCTGTACGTGGCCCCGTCAAAGGCCGAGCTCCAAGCGAGCCCTGCCAAGCCC TCGGGGAGAGCGGCCCGGACTCCATGATCCCGAGGAGGAGGACGATGAAGACGACGAGGAC GCGGGGACGGGCCGGAACGCTGGGGAACCTGTCTACGCGGACCTGATCACCCGCGCC ATCGAGAGCTCCCCGACAAACGGCTCACTCTGTCCAGATCTACGAGTGGATGGTGCCTTGC GTGCCCTACTTCAAGGATAAGGGCGACAGCAACAGCTCTGCCGGCTGGAAGAACTCCATCCGG CACAACCTGTCACTGCATAGTCGATTTCATGCGGGTCCAGAATGAGGGAACCTGGCAAGAGCTCT TGGTGGATCATCAACCTGATGGGGGAAGAGCGGAAAAGCCCCCGCGGGCGGGCTGTCTCC ATGGACAATAGCAACAAGTATACCAAGAGCCGTGGCCGCGCAGCCAAGAAGAAGGCAAGCCCTG CAGACAGCCCCGAATCAGCTGACGACAGTCCCTCCAGCTCTCCAAGTGGCCTGGCAGCCCC ACGTCACGCAGCAGTATGAGCTGGATGCGTGGACGGAATTCGTTACGCACCAATTCTAAC GCCAGCACAGTCACTGGCCGCTGTGCGCCATCATGGCAAGCACAGAGTTGGATGAAGTCCAG GACGATGATGCGCTCTCTCGCCATGCTCTACAGCAGCTCAGCCAGCCTGTCACTTTCAGTA AGCAAGCCGTGCACGGTGGAACTGCCACGGCTGACTGATATGGCAGGCACCATGAATCTGAAT GATGGGCTGACTGAAAACCTCATGGACGACCTGTGGATAACATCAGCTCCCGCCATCCACG CCATCGCCCACTGGGGGACTCATGCAGCGGAGCTCTAGCTTCCCGTATACCACCAAGGGCTCG GGCTGGGCTCCCCAACAGCTCCTTAAACAGCACGGTGTTCGGACCTTCATCTCTGAAGTCC CTACGCCAGTCTCCCATGCAGACCATCCAAGAGAACAAGCCAGCTACCTTCTCTCCATGTCA CACTATGGTAACAGACACTCCAGGACCTGCTCACTTCGGACTCACTTAGCCACAGCGATGTC ATGATGACACAGTCCGACCCCTTGATGTCTCAGGCCAGCACCGCTGTGTCTGCCAGAAATTC CGCCGGAACGTGATGCTTCGCAATGATCCGATGATGTCTTTGCTGCCAGCCTAACAGGGA AGTTTGGTCAATCAGAACTTGCTCCACCACCAGCACCAAACCCAGGGCGCTCTTGGTGGCAGC CGTGCCCTGTGCAATTCTGTGAGCAACATGGGCTTGAGTGAAGTCCAGCAGCCTTGGGTACGCC AAACACCAGCAGCTCTCTGTGAGCCAGTCTATGCAACCCCTCTCGGACTCTCTCTCAGGC TCCTCCTTGTAAGTCACTAGTGCAAACTGCCCCGTATGGGCGATGAGAAGTTCCCCAGCGAC TTGGACCTGGACATGTTCAATGGGAGCTTGAATGTGACATGGAGTCCATTATCCGTAGTGAA CTCATGGATGCTGATGGGTGGATTAACTTTGATTCCCTCATCTCCACACAGAATGTGTT GGTTTGAACGTGGGGAACCTCACTGGTGTAAAGCAGGCCTCATCTCAGAGCTGGGTGCCAGGC TGAAGGATCACTGAGGAAGGGGAAGTGGGCAAAGCAGACCCCTCAAACCTGACACAAGACCTACA GAGAAAACCCCTTGCCAAATCTGCTCTCAGCAAGTGGACAGTGATACCGTTTACAGCTTAACA CCTTTGTGAATCCCACGCCATTTCTTAACCCAGCAGAGACTGTTAATGGCCCTTACCCTGG GTGAAGCACTTACCCTTGGAAACAGAAGCTCTAAAAGTATGCAAAATCTTCC		
	ORF Start: ATG at 49		ORF Stop: TGA at 1828
	SEQ ID NO: 58	593 aa	MW at 63891.6kD
NOV16a, CG110410-01 Protein Sequence	MAEAPASPAPLSPLEVLDFEFEPQSRPRCTWPLQRPQLQASPAPKPSGETAADSMIPEEEDD EDDDDGGGRAGNAGNLSYADLITRAIESPDKRLTLSQIYEWVRCVPYFKDKGDSNSSAGW KNSIRHNLHLHFRMRVQNEGTGKSSWWIINPDGGKSGKAPRRRAVSMDSNKNYTKSRGRAAK KKAALQTAPESADDSPLSKWPGSPTSRSSDELDAWDFRSRTNSNASTVSGRLSPIMASTE LDEVQDDAPLSPMLYSSSASLSPSVSKPCTVELPRLTDMAGTMNLNDGLTENLMDLLDNIT LPPSQPSPTGGLMQRSSFFPYTTKSGSLGSPSTSSFNSTVFGPSSLNLSRQSPMQTIQENKPAT FSSMSHYGNQTLQDLLTSDLSHSDVMMTQSDPLMSQASTAVSAQNSRRNVMLRNDPMMSFAA QPNQGSVLNQNLHHQHQTQALGGSRLSNSVSNMGLSESSSLGSAKHQQQSPVQSQMOTLS DSLSGSSLYSTANLPMVGHEKFPDLDLDMFNGSLECDMESIIRSELMADGLDFNFDLSLIS TQNVVGLNVGNFTGAKQASSQSWVPG		

5

Further analysis of the NOV16a protein yielded the following properties shown in Table 16B.

Table 16B. Protein Sequence Properties NOV16a

PSort analysis:	0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV16a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 16C.

5

Table 16C. Geneseq Results for NOV16a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y96449	Forkhead transcription factor FKHL1 - Homo sapiens, 673 aa. [WO200031291-A1, 02-JUN-2000]	75..593 155..673	519/519 (100%) 519/519 (100%)	0.0
AAB56951	Human prostate cancer antigen protein sequence SEQ ID NO:1529 - Homo sapiens, 233 aa. [WO200055174-A1, 21-SEP-2000]	367..593 7..233	227/227 (100%) 227/227 (100%)	e-126
AA Y96448	Forkhead transcription factor FKHR - Homo sapiens, 655 aa. [WO200031291-A1, 02-JUN-2000]	75..593 158..655	256/530 (48%) 341/530 (64%)	e-120
AAB06076	Human homologue of Caenorhabditis elegans DAF-16 - Homo sapiens, 655 aa. [WO200033068-A1, 08-JUN-2000]	75..593 158..655	256/530 (48%) 341/530 (64%)	e-120
ABG20865	Novel human diagnostic protein #20856 - Homo sapiens, 837 aa. [WO200175067-A2, 11-OCT-2001]	127..593 392..837	208/478 (43%) 291/478 (60%)	3e-90

In a BLAST search of public sequence databases, the NOV16a protein was found to have homology to the proteins shown in the BLASTP data in Table 16D.

Table 16D. Public BLASTP Results for NOV16a

Protein Accession Number	Protein/Organism/Length	NOV16a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
--------------------------	-------------------------	---------------------------------	--	--------------

O43524	Forkhead box protein O3A (Forkhead in rhabdomyosarcoma-like 1) (AF6q21 protein) - Homo sapiens (Human), 673 aa.	75..593 155..673	519/519 (100%) 519/519 (100%)	0.0
Q9WVH4	Forkhead protein FKHR2 - Mus musculus (Mouse), 672 aa.	39..593 106..672	506/567 (89%) 523/567 (91%)	0.0
Q9BZ04	BA653O20.1 (forkhead box O3A (forkhead Drosophila homolog like 1, FKHL1)) - Homo sapiens (Human), 484 aa.	127..593 18..484	466/467 (99%) 467/467 (99%)	0.0
Q90YK2	Forkhead protein xFKHR1 - Xiphophorus maculatus (Southern platyfish), 664 aa.	1..593 1..664	367/684 (53%) 439/684 (63%)	e-176
Q9W7F8	Forkhead protein FKHR - Brachydanio rerio (Zebrafish) (Zebra danio), 651 aa.	10..593 8..651	370/662 (55%) 443/662 (66%)	e-172

Pfam analysis predicts that the NOV16a protein contains the domains shown in the Table 16E.

Table 16E. Domain Analysis of NOV16a			
Pfam Domain	NOV16a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Fork_head	46..159	39/117 (33%) 84/117 (72%)	2.9e-20

5

Example 17.

The NOV17 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 17A.

Table 17A. NOV17 Sequence Analysis			
	SEQ ID NO: 59	494 bp	
NOV17a. CG110882-01 DNA Sequence	AACGCGGCACATGGGAAGGAGCTCTACACACAGTGCCAGGCCTGCCACAAGCCCGTTGAAAAC TTCGTCGGCCCGAAGCACTGTGGCCTCATCGGCCGTCCAGCAGCCAGTGTGCCGGGATACGAC TATTCAGAAGGCATGAAGGCGTCGGGACTCACCTGGGACGAATCGACGCTCGATCAATTCTC ACTTCGCCGCTAGCCTTCGTCATGGCAGCAAGATGGGTTTTGCCGGATTGCGATAACCCGAGT GACCGGGCCGATGTCATTGCCTGGCTGCCGAAGATGAATGACGATCCACCATCTGCCGAAG AAGAGCTGACACCCATGCGCAGCAACGCTCATGCGAGCTGCCTCCCGCGTTGCCTGTGGCG CCTTGCTCGTCATGTCCGCCGCGCATGCGCGGCGCGACACCGTCTGACCCGCGCTCGATCG GCGGCGGCGAATGCGCCAAGAATGCTTATAACTGTGTGGGTGCCGCCAGGACG		
	ORF Start: at 1		ORF Stop: TGA at 322
	SEQ ID NO: 60	107 aa	MW at 11607.0kD

NOV17a, CG110882-01 Protein Sequence	NAAHGKELYTQCQACHKPVENFVGPKHCGLIGRPAASVPGYDYSEGMKASGLTWDESTLDQFL TSPVAFVNGTKMGFAGFDNPSDRADVIAWLRKMNDPTICPKKS
--	--

Further analysis of the NOV17a protein yielded the following properties shown in Table 17B.

Table 17B. Protein Sequence Properties NOV17a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.2852 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV17a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 17C.

Table 17C. Geneseq Results for NOV17a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB71120	Drosophila melanogaster polypeptide SEQ ID NO 40152 - Drosophila melanogaster, 108 aa. [WO200171042-A2, 27-SEP-2001]	5..94 11..104	40/94 (42%) 59/94 (62%)	2e-15
AAG38073	Arabidopsis thaliana protein fragment SEQ ID NO: 46915 - Arabidopsis thaliana, 112 aa. [EP1033405-A2, 06-SEP-2000]	1..95 11..109	39/99 (39%) 58/99 (58%)	8e-15
AAG16602	Arabidopsis thaliana protein fragment SEQ ID NO: 17311 - Arabidopsis thaliana, 112 aa. [EP1033405-A2, 06-SEP-2000]	1..95 11..109	39/99 (39%) 58/99 (58%)	8e-15
AAG27047	Zea mays protein fragment SEQ ID NO: 31733 - Zea mays subsp. mays, 119 aa. [EP1033405-A2, 06-SEP-2000]	1..99 11..113	39/103 (37%) 58/103 (55%)	2e-14
AAY77943	A. thaliana environmental stress tolerance related protein - Arabidopsis thaliana, 112 aa. [WO200008187-A2, 17-FEB-2000]	1..95 11..109	40/99 (40%) 57/99 (57%)	3e-14

10

In a BLAST search of public sequence databases, the NOV17a protein was found to have homology to the proteins shown in the BLASTP data in Table 17D.

Table 17D. Public BLASTP Results for NOV17a				
Protein Accession Number	Protein/Organism/Length	NOV17a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P00082	Cytochrome C2 - Rhodomicrobium vannielii, 104 aa.	5..97 7..102	47/96 (48%) 70/96 (71%)	6e-22
Q8YEH5	Cytochrome C-552 - Brucella melitensis, 202 aa.	3..104 75..182	53/108 (49%) 71/108 (65%)	2e-21
Q939T7	Flavocytochrome c cytochrome subunit - Rhodovulum sulfidophilum (Rhodobacter sulfidophilus), 238 aa.	5..105 34..140	43/107 (40%) 67/107 (62%)	4e-17
Q53144	ISOCYTOCHROME C2 precursor - Rhodobacter sphaeroides (Rhodopseudomonas sphaeroides), 144 aa.	3..101 26..130	43/105 (40%) 63/105 (59%)	9e-17
Q98BN4	Cytochrome c - Rhizobium loti (Mesorhizobium loti), 215 aa.	1..104 71..181	49/111 (44%) 63/111 (56%)	1e-16

- 5 Pfam analysis predicts that the NOV17a protein contains the domains shown in the Table 17E.

Table 17E. Domain Analysis of NOV17a			
Pfam Domain	NOV17a Match Region	Identities/ Similarities for the Matched Region	Expect Value
cytochrome_c	1..97	41/116 (35%) 76/116 (66%)	7e-23

Example 18.

- 10 The NOV18 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 18A.

Table 18A. NOV18 Sequence Analysis			
	SEQ ID NO: 61	4529 bp	
NOV18a, CG111188-01 DNA	GCGCCGCGTCTTCCCGGTCTCCTTTCCCGGCCGCACAGGGTTTATAGGATCACATTGACAAA AGTACCATGGAGTTTATGAGTCAGCATATTTATTGTTCTTATTCCTCAATAGTTATTACA		

Sequence	<p> GTAATTTTCCTCTTCTCTGGCTTTTCATGAAAGAAACATTATATGATGAAGTTCTTGCAAAA CAGAAAAGAGAACAAAAGCTTATTCTTACCAAAACAGATAAAAAGAAAGCAGAAAAGAAAAAG AATAAAAAGAAAGAAATCCAGAATGGAAACCTCCATGAATCCGACTCTGAGAGTGTACCTCGA GACTTTAAATTATCAGATGCTTTGGCAGTAGAAGATGATCAAGTTGCACCTGTTCATTGAAT GTCGTTGAAACTTCAAGTAGTGTAGGGAAAGAAAAAGAAAGGAAAAAGCAAAAAGCCTGTG CTTGAAGAGCAGGTCAATCAAGAAAGTGACGCATCAAGATTCTTGGCAAAAAGTAGAACCT GTCCAGTTACTAAACAGCCCAACCTCCCTCTGAAGCAGCTGCCTCGAAGAAGAAACCAGGG CAGAAGAAGTCTAAAAATGGAAGCGATGACCAGGATAAAAAGGTGGAACTCTCATGGTACCA TCAAAAAGGCAAGAAGCATTGCCCTCCACCAAGAGACTAAACAAGAAAGTGGATCAGGGAAG AAAGCTTCATCAAGAAACAAAAGACAGAAAATGTCTTCGTAGATGAACCCCTTATTATGCA ACTACTTATATTCCTTTGATGGATAATGCTGACTCAAGTCCTGTGGTAGATAAGAGAGAGGTT ATTGATTGTCTTAAACCTGACCAAGTAGAAGGGATCCAGAAATCTGGGACTAAAAAAGTGAAG ACCGAACTGACAAAGAAATGCTGAAGTGAAGTTTAAAGATTTCTTCTGTCTTGAAGACT ATGATGTTTTCTGAAGATGAGGCTCTTTGTGTGTAGACTTGCTAAAGGAGAAGTCTGGTGTA ATACAAGATGCTTTAAAGAAGTCAAGTAAGGAGAATGACTACGCTTATACATCAGCTTCAA GAAAAGGACAAGTTACTCGCTGCTGTGAAGGAAGATGCTGCTCTACAAAGGATCGGTGTAAAG CAGTTAACCCAGGAAATGATGACAGAGAAAGAAAGCAATGTGGTTATGACAAGGATGAAA GATCGGATTGGAACATTAGAAAAGGAACATAATGTATTTCAAAACAAAATACATGTCACTAT CAAGAGACTCAACAGATGCAGATGAAGTTTCAGCAAGTTCGTGAGCAGATGGAGGCAGAGATA GCTCACTTGAAGCAGGAAAATGGTATACTGAGAGATGCAGTCAGCAACACTACAAATCAACTG GAAAGCAAGCAGTCTGCAGAACTAAATAAACTACGCCAGGATTATGTAGGTTGGTGAATGAG CTGACTGAGAAAACAGGAAAGCTACAGCAAGAGGAAGTCCAAAAGAAGATGCTGAGCAAGCA GCTACTCAGTTGAAGTTCAACTACAAGAAGCTGAGAGAAGGTGGGAAGAAGTTCAGAGCTAC ATCAGGAAGAGAACAGCGGAACATGAGGCAGCACAGCAAGATTTACAGAGTAAATTTGTGGCC AAAGAAAATGAAGTACAGAGTCTGCATAGTAAGCTTACAGATACCTTGGTATCAAAACACAG TTGGAGCAAAGACTAATGCAGTTAATGGAATCAGAGCAGAAAAGGTTGAACAAAGAAGAGTCT CTACAAATGCAGGTTCAAGATATTTTGGAGCAGAATGAGGCTTTGAAAGCTCAAAATTCAGAC TTCCATTCCCAGATAGCAGCCAGACCTCCGCTTCAGTTCTAGCAGAAGAATTACATAAAGTG ATTGCAGAAAAGGATAAGCAGATAAAACAGACTGAAGATTCTTAGCAAGTGAACGTGATCGT TTAACAAGTAAAGAAGAGGAACTTAAGGATATACAGAATATGAATTTCTATTAAAAGCTGAA GTGCAGAAATTACAGGCCCTGGCAATGAGCAGGCTGCTGCTGCATGATGGAGAAGATG CAACAAAGTGTATTATGTTAAAGATGATAAAATAAGATTGCTGGAAGAGCAACTACAACATGAA ATTTCAAAACAAAATGGAAGAATTTAAGATTCTAAATGACCAAAACAAAGCATTAAATCAGAA GTTCAAGAGCTACAGACTCTTGTCTTGAACAGCCTAATAAGGATGTTGTGGAACAAATGGAA AAATGCATTCAAGAAAAGATGAGAAGTTAAAGACTGTGGAAGAATTACTTGAACCTGGACTT ATTCAGGTGGCAACTAAAGAAGAGGAGCTGAATGCAATAAGAACAGAAAATTCATCTTGACA AAAGAAGTTCAAGACTTAAAGCTAAGCAAAATGATCAGGTTCTTTTGCCTCTCTAGTTGAA GAACTTAAAGAAAGTATCCATGAGAAAGATGGAAGATCAAGTCTGTAGAAGAGCTTCTGGAG GCAGAACTTCTCAAAGTTGCTAACAAAGGAGAAAACGTTCAGGATTTGAAACAGGAAAATAAG GCTCTAAAAGAAGAAAATAGGAAATGTCCAGCTTGAAAAGGCTCAACAGTTATCTATCACTTCC AAAGTTCAAGGAGCTTCAGAACTTATTAAAAGGAAAAGAGGAACAGATGAATACCATGAAGGCT GTTTTGGAAGAGAAAGAGAAAGACCTAGCCAATACAGGGAAGTGGTTACAGGATCTTCAAGAA GAAAATGAATCTTTAAAGCACATGTTTCAAGGAAGTAGCACAACTAATGAAAGAGGCTCT TCTGCATCACAGTTGAAGAACTTGAAGATTGTGTTGAAAGAAAAGGGAATGAATTGAAGAGG TTAGAAGCCATGTCTAAAAGAGAGGAGAGTGATCTTTTACGAAAACACAGCTGTTACAGGAT GTACAAGATGAAAACAAATTGTTTAAAGTCCCAATTTAGCAGCTTAAACAACAAAACCTACCAA CAGGCATCTTCTTTCCCTCATGAAGAATTATTAAGTAATTTTCAAGAGAGAGAAAGAA ATAAGTGGTCTCTGGAATGAGTTAGATTCTTTGAAGGATGCAGTTGAACACCAGAGGAAGAAA AACAAATGACCTTCGGGAGAAAACTGGGAAGCAATGGAAGCATTGGCATCAACTGAAAAATG CTGCAGGACAAAGTGAACAAGACTTCCAAGGAAAGGCAGCAACAGGTGGAAGCTGTTGAGTTG GAGGCTAAAGAAGTTCTCAAAAATATTTCCAAAGGTGTCTGTCCCTTCTAATTTGAGTTAT GGTGAATGGTTGCATGGATTGAAAAAAGGCAAAAGAATGTATGGCTGGAACCTTCAAGGTCA GAGGAGTTAAGGTTCTAGAGCACAAGTTGAAAGAAGCTGATGAAATGCACATGTTTACAG CTAGAGTGTGAAAAATACAAATCCGTCTTGCAGAAAACAGAAAGGAATTTTACAGAACTACAG AGAAGTGTGAGCAAGAAGAAAAATAATGGAAGTTAAGTTCGATGAATCACACAAGACTATT AAACAGATGCAGTCATATTTACATCTTCAAGAAAGAGCTAGAGCGATTAGAAGCGAAAAAT AAGGATATTGAAAATCTGAGAAGAGAACGAGAACATTTGGAATGGAAGTGAAGGAGGAGAG ATGGAACGATCTACCTATGTTACAGAACTCAGAGAGTTGAAGGCACAGTTAAATGAAACACTC ACAAAACCTAGAACTGAACAAAATGAAAGACAGAAAGGTAGCTGGTGATTGTCATAAGGCTCAA CAGTCACTGGAGCTTATCCAGTCAAAAATAGTAAAGCTGCTGGAGACACTACTGTTATTGAA AATAGTGATGTTTTCCCGAGAAACGGAGTCTTCTGAGAAGGAGACAATGTCTGTAAGTCTAAAT CAGACTGTAACACAGTTACAGCAGTTGCTTTCAGGCGGTAACCAACAGCTCACAAGGAGAAA GAGCACTACCAGGTGTAGAGTGAAGTAATTGGGAACTGTTTCAATTTGAGGATAAAAAAGGCA TTGTATTATATTTGCCAAATTAAGCCTTATTTATGTTTTACCCTTTCTACTTTGTCAGAA ACACTGAACAGAGTTTGTCTTTCTAATCCTTGTAGACTACTGATTTAAAGAGGAAAAAA AAAAGCCAACCTCTGTAGACACCTCAGAGTTTGTATTTATAATAAAACTGTTTGAATAATTA </p>
----------	---

	GACCTTTACATTCTCTGAAGATAAACATGTAATCTTTTATCTTATTTTGCTCAATAAAATTGTT CAGAAGATCAAAGTGGTAAAGACAATGTAATTTTAACATTTTAATACTGATGTTGTACACTG TTTTACTTAACATTTTGGGAAGTAACTGCCTCTGACTTCAACTCAAGAAACACTTTTTGTT GCTAATGTAATCGGTTTTTGTAAATGGCGTCAGCAAATAAAGGATGCTTATTATTC		
	ORF Start: ATG at 70		ORF Stop: TGA at 4054
	SEQ ID NO: 62	1328 aa	MW at 152804.3kD
NOV18a, CG111188-01 Protein Sequence	MEFYESAYFIVLIPSIVITVIFLFFWLFMKETLYDEVLAQKREQKLIPTKTDKKKAEKKKNK KKEIQNGNLHESDSESVPDFKLSDALAVEDDQVAPVPLNVVETSSSVRERKKKKKQKPVLE EQVIKESDASKIPGKKVEPVPTKQPTPPSEAAASKKKPGQKKSKNGSDQDKKVTLMVPSK RQEALPLHQETKQESGSGKKASSKKQKTENVFVDEPLIHATTYIPLMDNADSSPVVDKREVID LLKPDQVEGIQKSGTKKLKTETDKENAIEVKFKDFLLSLKTMFSEDEALCVVDLLKEKSGVIO DALKKSSKGELTTLIHQLEKDKLLAAVKEDAAATKDRCKQLTQEMMTEKERSNVMTRMKDR IGTLEKEHNVFQNKIHVSQYQETQMQMKFQQVREQMEAEIAHLKQENGILRDAVSNTTNQLES KQSAELNKLQDYARLVNELTEKTKLQEEVQKKNAEQAAATQLKVQLQEAERRWEESYIR KRTAEHEAAQQDLQSKFVAKENEVQSLHSLKLTDTLVSKQQLQRLMQLMESEQKRVNKEESLQ MQVDILEQNEALKAQIQFHSQIAAQTASVLAELHKVIAEKDKQIKQTEDSLASERDRLT SKEELKDIQNMNFFLLKAEVQKLQALANEQAAAAHELEKMQQSVYVKKDKIRLLEQLQHEIS NKMEEFKILNDQNKALKSEVQKLQTLVSEQPNKDVVEQMEKCIQEKDEKLKTVEELETLIQ VATKEEELNAIRTESSSLTKEVQDLKAKQNDQVSFASLVEELKKVIEHKDGKIKSVBELLAE LLKVANKKTVQDLKQEI KALKEEIGNVQLEKAQQLSITSKVQELQNLKGGKEQMNTMKAVL EEKEKDLANTGKWLQDLQEEENSLKAHVQEVQAHNLKEASSAQFEELIVLKEKGNELKRL AMLKERESDLSSKTQLLDVQDENKLFKSQIEQLKQONYQOASSFPHEELLKVISEREKEIS GLWNELDSLKDAVEHQKKNDLREKNWEAMEALASTEKMLODKVNKTSKERQQQVEAVELEA KEVLKFLPKVSVPSNLSYGEWLHGFKEKAKCECMAGTSGSEEVKLEHLKEADEMHTLLQLE CEKYKSVLAETEGILQKLQRSVEQEENKWKVKVDESHKTIKQMSSFTSSEQELERLRSENKD IENLRREHLEMELEKAEMERSTYVTEVRELKAQLNETLTKLRTEQNERQKVAGDLHKAQQS LELIQSKIKAAGDTTVIENSVDSPETESSEKETMSVSLNQTVTQLQQLQAVNQQLTKEKEH YQVLE		
	SEQ ID NO: 63	4326 bp	
NOV18b, CG111188-02 DNA Sequence	GCCGCACAGGGTTTTATAGGATCACATTGACAAAAGTACCATGGAGTTTTATGAGTCAGCATA TTTTATTGTTCTTATCTCTCAATAGTTATTACAGTAATTTTCTCTCTCTGCTTTTCAT GAAAGAAACATTATATGATGAAGTTCTTGCAAAACAGAAAAGAGAACAAGCTTATTCCTAC CAAAACAGATAAAAAAGAAAGCAGAAAAGAAAAAGAAATAAAAGAAAGAAATCCAGAATGGAAA CCTCCATGAATCCGACTCTGAGAGTGACCTCGAGACTTTAAATTATCAGATGCTTTGGCAGT AGAAGATGATCAAGTTGCACTGTTCCATTGAATGTCGTTGAACTTCAAGTAGTGTAGGGA AAGAAAAAGAAAGGAAAAGAAACAAAAGCCTGTGCTTGAAGAGCAGGTCATCAAGAAAGTGA CGCATCAAGATTCTGGCAAAAAGTAGAACCTGTCCCAGTTACTAAACAGCCACCCCTCC CTCTGAAGCAGCTGCCTCGAAGAAGAAACAGGGCAGAGAAGTCTAAAAATGGAAGCGATGA CCAGGATAAAAAGTGGAAGTCTCATGGTACCATCAAAAAGGCAAGAGCATTGCCCTCCA CCAAGAGACTAAACAGAAGTGGATCAGGGAAGAAAGAGCTTATCAAGAAACAAAGAC AGAAAATGTCTTCGTAGATGAACCCCTTATTATGCACTACTTATATCTTTTGTAGGATAA TGCTGACTCAAGTCTGTGTAGATAAGAGAGAGGTTATTGATTGTGTTAAACCTGACCAAGT AGAAGGGATCCAGAAATCTGGGACTAAAAAAGTGAAGACCGAACTGACAAAGAAAATGCTGA AGTGAAGTTTAAAGATTTCTTGTCTGCTTGAAGACTATGATGTTTCTGAAGATGAGGCTCT TTGTGTTGTAGACTTGCTAAAGGAGAAGTCTGGTGAATACAAGATGCTTTAAAGAAGTCAAG TAAGGGAGAATTGACTACGCTTATACATCAGCTTCAAGAAAAGGACAAGTTACTCGCTGCTGT GAAGGAAGATGCTGCTGCTACAAAGGATCGGTGTAAGCAGTTAACCCAGGAAATGATGACAGA GAAAGAAAGAAAGCAATGTGTTTATAACAAGGATGAAAGATCGAATTGGAACATTAGAAAAGGA ACATAATGTATTTCAAAACAAAATACATGTCAGTTATCAAGAGACTCAACAGATGCAGATGAA GTTTCAGCAAGTTCTGTAGCAGATGGAGGCAGAGATAGCTCACTTGAAGCAGGAAAATGGTAT ACTGAGAGATGCAGTCAGCAACACTACAAATCAACTGGAAAGCAAGCAGTCTGCAGAACTAAA TAACTACGCCAGGATTATGCTAGGTGTTGTAATGAGCTGACTGAGAAAAAGGAAAGCTACA GCAAGAGGAAGTCCAAAAGAAAGTGTGAGCAAGCAGCTACTCAGTTGAAGTTCAACTACA AGAAGCTGAGAGAAGGTGGGAAGAGTTTCAAGCTACATCAGGAAGAGAACAGCGGAACATGA GGCAGCACAGCAAGATTTACAGAGTAAATTTGTGGCCAAAGAAAATGAAGTACAGAGTCTGCA TAGTAAGCTTACAGATACCTTGGTATCAAAACACAGTTGGAGCAAGAGCTAATGCAGTTAAT GGAATCAGAGCAGAAAAGGTGAACAAAGAGAGTCTCTACAAATGCAGGTTCCAGATATTTT GGAGCAGAAATGAGGCTTTGAAAGCTCAAAATTCAGCAGTTCCATTCAGATAGCAGCCAGAC CTCCGCTTCAAGTTCTAGCAGAAAGTATACATAAAGTGATTGCAGAAAAGGATAAGCAGATAAA ACAGACTGAAGATTCTTTAGCAAGTGAACGTGATCGTTTAAACAGTAAAGAGAGGAACCTTAA GGATATACAGAAATATGAATTTCTTATTAAGCTGAAGTGCAGAAATACAGGCCCTGGCAAA TGAGCAGGCTGCTGCTGCATGAATTTGGAGAAGATGCAACAAAGTGTATGTTTAAAGATGA TAAATAAGATTGCTGGAAGAGCAACTACAACATGAAATTTCAACAAAATGGAAGAATTTAA GATTCTAAATGACCAAAACAAAGCATTAAATCAGAAGTTCAGAAGCTACAGACTCTTGTTC		

	TGAACAGCCTAATAAGGATGTTGTGGAACAAATGGAAAAATGCATTCAAGAAAAAGATGAGAA GTTAAAGACTGTGGAAGAATTACTTGAAGCTGGACTTATTCAGGTGGCACTAAAGAAGAGGA GCTGAATGCAATAAGAACAGAAAAATTCATCTCTGACAAAAAGATTCAAGACTTAAAGCTAA GCAAAATGATCAGGTTCTTTTGCCTCTCTAGTTGAAGAACTTAAGAAAGTGATCCATGAGAA AGATGGAAGATCAAGTCTGTAGAAGAGCTTCTGGAGGCAGAACTTCTCAAGTTGCTAACAA GGAGAAACTGTTCAGGATTGAAACAGGAAATAAAGGCTCTAAAAAGAGAAATAGGAAATGT CCAGCTTGAAGAGGCTCAACAGTTATCTATCACTTCCAAAGTTCAGGAGCTTCAGAACTTATT AAAAGGAAAAAGAGGAACAGATGAATACCATGAAGGCTGTTTTGGAAGAGAAAGAGAAAGACCT AGCCAATACAGGGAAGTGGTTACAGGATCTTCAAGAAGAAATGAATCTTTAAAGCACATGT TCAGGAAGTAGCACAACTAATCTGAAAGAGGCTCTTCTGCATCACAGTTGAAGAACTTGA GATTGTGTTGAAAGAAAGGAAATGAATTGAAGAGGTTAGAAGCCATGCTAAAAGAGAGGGA GAGTGATCTTTCTAGCAAAACACAGCTGTTACAGGATGTACAAGATGAAACAAATTTGTTAA GTCCCAATTGAGCAGCTTAAACAACAACTACCAACAGGCATCTCTTTTCCCCCTCATGA AGAATTATTAAGTAATTTAGAAAGAGAGAGAAAGAAATAAGTGGTCTCTGGAATGAGTTAGA TTCTTTGAAGGATGCAGTTGAACACCAGAGGAAGAAAAACAATAGTTATGGTGAATGGTTGCA TGGATTTGAAAAAAGGCAAGAAATGTATGGCTGGAACCTCAGGGTCAGAGGAGGTTAAGGT TCTAGAGCACAAGTTGAAAGAAGCTGATGAAATGCACACATTGTTACAGCTAGAGTGTAAGAA ATACAAATCCGTCCTTGCAAAACAGAAGGAATTTTACAGAAGCTACAGAGAAGTTGAGCA AGAAGAAATAAATGGAAGTTAAGGTCGATGAATCACACAAGACTATTAAACAGATGCAGTC ATCATTTACATCTTCAGAACAGAGCTAGAGCGATTAAGAAGCGAAAAATAAGGATATTGAAAA TCTGAGAAGAGAACGAGAACATTTGGAATGGAAGTGAAGAGGAGAGATGGAACGATCTAC CTATGTTACAGAAGTCAGAGAGTTGAAGGCACAGTTAAATGAAACACTCACAACTTAGAAC TGAACAAATGAAAGACAGAAGGTAGCTGGTGATTGTCATAAGGCTCAACAGCTCAGGAGCT TATCCAGTCAAAATAGTAAAGCTGCTGGAGACACTACTGTTATTGAAATAGTGATGTTTC CCCAGAAACGGAGTCTTCTGAGAAGGAGACATGTCTGTAAGTCTAAATCAGACTGTAACACA GTTACAGCAGTTGCTTCAGGCGGTAAACCAACAGCTCACAAGGAGAGAGAGCACTACCAGGT GTTAGAGTGAAGTAATTGGGAACTGTTTATTGAGGATAAAAAAGGCATTGTTATATTTT GCCAAATTAAAGCCTTATTATGTTTTCACCTTTTCTACTTTGTCAGAAACACTGAACAGAGT TTTGTCTTTTCTAATCCTTGTAGACTACTGATTTAAAGAAGGAAAAAAGGCCAATCTG TAGACACCTTCAGAGTTTAGTTTATAATAAACTGTTTGAATAATTAGACCTTTACATTCC TGAAGATAAACATGTAATCTTTATCTTATTTGCTCAATAAAATGTTTGAAGATCAAGT GGTAAAGACAATGTAAATTAACATTTTAATACTGATGTTGTACACTGTTTACTTAACTT TTGGGAAGTAACTGCCTCTGACTTCAACTCAAGAAAAACCTTTTTGTTGCTAATGTAATCGG TTTTGTAATGGCGTCAGCAATAAAGGATGCTTATTATTC		
	ORF Start: ATG at 41		ORF Stop: TGA at 3851
	SEQ ID NO: 64	1270 aa	MW at 146190.7kD
NOV18b, CG111188-02 Protein Sequence	MEFYESAYFIVLIPSIVITVIFLFFWLFMKETLYDEVLAQKQREKLIPTKTDKKKAEEKKNK KKEIQNGLHESDSVPRDFKLSDALAVEDDQVAPVPLNVVETSSSVRERKKKQKQKPVLE EQVIKESDASKIPGKKEVPVPTKQPTPPSEAAASKKKPGQKSKNGSDDQDKKVTLMVPSK RQEALPLHQETKQESGSGKKKASSKKQKTENVFVDEPLIHATTYIPLMDNADSSPVVDKREVI DLLKPDQVEGIQKSGTKLKTETDKENAEVKFKDFLLSLKTMFSEDEALCVVDLLKEKSGVI QDALKKSSKGLTTLIHQLQEKDKLLAAVKEDAAATKDRCKQLTQEMMTEKERSNVVITRMKD RIGTLEKEHNVFNKIHVSQYQETQOMQMKFQQVREQMEAEIAHLKQENGILRDAVSNTNQLE SKQSAELNKLQDYARLVNELTEKTKGLQEEVQKNAEQAATQLKVQLQEAEERWEEVQSYI RKRTAEHEAAQDLQSKFVAKENEVQSLHSLKLDTLVSKQLEQLRMLQMLESEQRVNKEESL QMQVQDILEQNEALKAQIQFHSQIAAQTASVLAELHKVIAEKDKQIKQETEDSLASERDRL TSKEEELKDIQNMNFKLKAQVQKLQALANEQAAAAHELEKMQQSVYVKKDKIRLLEEQLQHEI SNKMEEFKILNDQNKALKSEVQKLQTLVSEQPNKDVVEQMEKCIQEKDEKLKTVEELLETGLI QVATKEEELNARTENSSLTKEVQDLKAKQNDQVSFASLVEELKKVIHEKDGKIKSVEELLEA ELLKVANKEKTVQDLKQEIKAKEEIGNVQLEKAQOLSITSKVQELQNLKKEEQMNTMKAV LEEKEKDLANTGKWLQDLQEEENESLKAHVQEVQHNLEKASSASQFEELEIVLKEKENELKRL EAMLKERESDLSKTQLQDVQDENKLFKSQIEQLKQNYQOASSFPHEELLKVISEREKEI SGLWNELESLKDAVEHQKKNNSYGEWLHGFKEKAKECMAGTSGSEEVKVLEHLKEADEMHT LLQLECEKYKSVLAETEGILQKLQRSVEQEENKWKVDESHKTIKQMQSSPTSSEQLERLR SENKDIEENLRREHLEMELEKAEMERSTYVTEVRELKAQLNETLTKLRTEQNERQKVAGDLH KAQQSLELIQSKIKAAGDTTVIENS DVSPETESSEKETMSVSLNQTVTLQQLLQAVNQQLT KEKEHYQVLE		
	SEQ ID NO: 65	4416 bp	
NOV18c, CG111188-03 DNA Sequence	GCCGCACAGGGTTTATAGGATCACATTGACAAAAGTACCATGGAGTTTATGAGTCAGCATA TTTTATTGTTCTTATTCCTTCAATAGTTATTACAGTAATTTCTCTTCTCTCGGCTTTTCAT GAAAGAAACATTATATGATGAAGTCTTGCAAAACAGAAAAGAGAAACAAAGCTTATTCCTAC CAAAACAGATAAAAAGAAAGCAGAAAAGAAAAAGAAATAAAAAGAAAGAAATCAGAAATGGA CCTCCATGAATCCGACTCTGAGAGTGACCTCGAGACTTTAAATTATCAGATGCTTTGGCAGT AGAAGATGATCAAGTTGCACCTGTTCCATTGAATGTCGTTGAAACTTCAAGTAGTGTAGGGA		

	AAGAAAAAGAAGGAAAAAGAAACAAAAGCCTGTGCTTGAAGAGCAGGTCATCAAAGAAAGTGA CGCATCAAAGATTCTCTGGCAAAAAAGTAGAACCTGTCCCAGTTACTAAACAGCCCCCCTCC CTCTGAAGCAGCTGCCTCGAAGAAGAAACCAGGGCAGAGAAGTCTAAAAATGGAAGCGATGA CCAGGATAAAAAAGGTGGAACCTCTCATGGTACCATCAAAAAGGCAAGAAGCATTGCCCCCTCCA CCAAGAGACTAAACAAGAAAGTGGATCAGGGAAGAAGAAAGCTTCATCAAAGAAACAAAAGAC AGAAAATGTCTTCGTAGATGAACCCCTTATTTCATGCACTACTTATATTCCTTTTGATGGATAA TGCTGACTCAAGTCCTGTGGTAGATAAGAGAGAGGTTATTGATTGCTTAAACCTGACCAAGT AGAAGGGATCCAGAAATCTGGGACTAAAAAATGAAGACCGAAACTGACAAAGAAAATGCTGA AGTGAAGTTTAAAGATTTTCTTCTGTCTTGAAGACTATGATGTTTTCTGAAGATGAGGCTCT TTGTGTTGTAGACTTGCTAAAGGAGAAGTCTGGTGTAAATACAAGATGCTTTAAAGAAGTCAAG TAAGGGAGAATTGACTACGCTTATACATCAGCTTCAAGAAAAGGACAAGTTACTCGCTGCTGT GAAGGAAGATGCTGCTGTACAAAGGATCGGTGTAAGCAGTTAACCCAGGAAATGATGACAG GAAAGAAAGAAGCAATGTGGTTATAACAAGGATGAAAGATCGAATTGGAACATTAGAAAAGGA ACATAATGTATTTCAAACAAAATACATGTCAAGTTATCAAGAGACTCAACAGATGCAGATGAA GTTTCAGCAAGTTCGTGAGCAGATGGAGGCAGAGATAGCTCACTGAAGCAGGAAAATGGTAT ACTGAGAGATGCAGTCAGCAACACTACAAATCAACTGGAAGCAAGCAGTCTGCAGAACTAAA TAACTACGCCAGGATTATGCTAGGTTGGTGAATGAGCTGACTGAGAAAACAGGAAAGCTACA GCAAGAGGAAGTCCAAAAGAAGATGCTGAGCAAGCAGCTACTCAGTTGAAGGTTCAACTACA AGAAGCTGAGAGAAGGTGGGAAGGTTTCAAGCTACATCAGGAAGAGAACAGCGGAACATGA GGCAGCACAGCAAGATTTACAGAGTAAATTTGTGGCCAAAGAAAATGAAGTACAGAGTCTGCA TAGTAAGCTTACAGATACCTTGGTATCAAAACAACAGTTGGAGCAAAGACTAATGCAGTTAAT GGAATCAGAGCAGAAAAGGGTGAACAAAGAAGAGTCTCTACAATGCAGGTTCCAGGATATTTT GGAGCAGAATGAGGCTTTGAAAGCTCAAATTCAGCAGTTCATTCCAGATAGCAGCCAGAC CTCCGCTTCAGTTCTAGCAGAAGAATTACATAAAGTATTGTCAGAAAAGGATAAGCAGATAAA ACAGACTGAAGATTCTTTAGCAAGTGAACGTGATCGTTTAAACAAGTAAAGAAGAGGAACCTAA GGATATACAGAATATGAATTTCTTATTAAGGCTGAAGTGCAGAAATTACAGGCCCTGGCAAA TGAGCAGGCTGCTGCTGCACATGAATTGGAGAAGATGCAACAAAGTGTTTATGTTAAAGATGA TAAATAAGATTGCTGGAAGAGCAACTACAAATGAAATTTCAAACAAAATGGAGAATTAA GATTCTAAATGACCAAAACAAGCATTAAATCAGAAGTTCAGAAGCTACAGACTCTTGTTTC TGAACAGCCTAATAAGGATGTTGTGGAACAAATGGAAAATGCATTCAAGAAAAGATGAGAA GTTAAAGACTGTGGAAGAATTACTTGAAGCTGGACTTATTCAGGTGGCAACTAAAGAAGAGGA GCTGAATGCAATAAGAACAAGAAAATTCATCTCTGACAAAAGAAGTTCAAGACTTAAAGCTAA GCAAAATGATCAGGTTTCTTTTGCCTCTCTAGTTGAAGAAGTAAAGAAAGTGATCCATGAGAA AGATGGAAGATCAAGTCTGTAGAAGAGCTTCTGGAGGCAGAACTTCTCAAAGTTGCTAACAA GGAGAAAAGTGTTCAGGATTTGAAACAGGAAATAAAGGCTCTTAAAGAGAAATAGGAAATGT CCAGCTTGAAAAGGCTCAACAGTTATCTATCACTTCCAAGTTTCAGGAGCTTCAGAAGTATT AAAAGGAAAAGAGGAACAGATGAATACCATGAAGGCTGTTTTGGAAGAGAAAGAGAAAGCACT AGCCAATACAGGGAAGTGGTTACAGGATCTTCAAGAAGAAAATGAATCTTTAAAGCACATGT TCAGGAAGTAGCACACAATAACTTGAAAGAGGCCTCTCTGCATCACAGTTTGAAGAAGTGA GATTGTGTTGAAAGAAAAGGAAAATGAATTGAAGAGGTTAGAAGCCATGCTAAAAGAGAGGGA GAGTGATCTTTCTAGCAAAACACAGCTGTTACAGGATGTACAAGATGAAAACAAATGTTTAA GTCCCAAATTGAGCAGCTTAAACAACAAAACCTACCAACAGGCATCTTCTTTCCCTCATGA AGAATTATTAAAGTAATTTAGAAAAGAGAGAAAGAAATAAGTGGTCTCTGGAATGAGTTAGA TTCTTTGAAGGATGCAGTTGAACACCAGAGGAAGAAAACAAATGAAAGGCAGCAACAGGTGGA AGCTGTTGAGTTGGAGGCTAAAGAAGTTCTCAAAAAATTATTTCAAAGGTGCTGTGCCCTTC TAATTTGAGTTATGGTGAATGGTTGTCATGGATTGAAAAAAGGCAAAAGATGTATGGCTGG AACTTCAGGGTCAGAGGAGGTTAAGGTTCTAGAGCACAAGTTGAAAGAAGCTGATGAAATGCA CACATTGTTACAGCTAGAGTGTGAAAAATACAAATCCGCTCTTCAGAAAACAGAAGGAATTTT ACAGAAGCTACAGAGAAGTGTGAGCAAGAAGAAAATAAATGGAAGTTAAGGTGATGAATC ACACAAGACTATAAACAGATGCAGTCATCATTACATCTTCAGAACAGAGCTAGAGCGATT AAGAAGCGAAAATAAGGATATTGAAAATCTGAGAAGAGAACGAGAACATTGGAATGGAAC AGAAAAGGCAGAGATGGAACGATCTACCTATGTTACAGAAGTCAGAGAGTTGAAGGCACAGTT AAATGAAACACTCACAAAACCTAGAACTGAACAAAATGAAAGACAGAAGGTAGCTGGTGATTT GCATAAGGCTCAACAGTCACTGGAGCTTATCCAGTCAAAAAATAGTAAAGAGTCTGGAGACAC TACTGTTATTGAAAATAGTGATGTTTCCCAAGAACCGAGTCTTCTGAGAAGGAGACAAATGTC TGTAAGTCTAAATCAGACTGTAAACAGTTACAGCAGTTGCTTCAGGCGGTAACCAACAGCT CACAAAGGAGAAAGAGCACTACCAGGTGTTAGAGTGAAGTAATTGGGAACTGTTTCAATTTGAG GATAAAAAGGCATTGTATTATATTTTGCCAAATTAAAGCCTTATTTATGTTTTACCCCTTTC TACTTTGTGAGAAACACTGAACAGAGTTTTGTCTTTCTAATCCTTGTAGACTACTGATTTA AAGAAGGAAAAAAAAGCCAACTCTGTAGACACCTTCAGAGTTTAGTTTTATAATAAAAACT GTTTGAATAATTAGACCTTTACATTCCTGAAGATAAACATGTAATCTTTTATCTTATTTTGCT CAATAAAATTTGTTCAAGATCAAGTGGTAAAGACAATGTAATAATTTAACATTTTAACTG ATGTTGTACACTGTTTTACTTAACTTTTGGGAAGTAACCTGCTGACTTCAACTCAAGAAA ACACTTTTTGTGCTAATGTAATCGGTTTTTGTAAATGGCGTCAGCAAAATAAAGGATGCTTA TTATTC	
	ORF Start: ATG at 41	ORF Stop: TGA at 3941

	SEQ ID NO: 66	1300 aa	MW at 149609.6kD
NOV18c, CG111188-03 Protein Sequence	MEFYESAYFIVLIPSIVITVIFLFFWLFMKETLYDEVLAQKREQKLIPTKTDKKKAEKKKNK KKEIQNGNLHESDSSESVPRDFKLSDALAVEDDQVAPVPLNVVETSSSVRERKKKEKKQKPVLE EQVIKESDASKIPGKKVEPVPTKQPTPPSEAAASKKPGQKSKNGSDQDKKVVETLMVPSK RQEALPLHQETKQESGSGKKKASSKKQKTENVFVDEPLIHATTYIPLMDNADSSPVVDKREVI DLLKPDQVEGIQKSGTKKLTETDKENAEVFKDFLLSLKTMFSEDEALCVVDLLKEKSGVI QDALKKSSKGELTTLIHQLEKDKLLAAVKEDAAATKDRCKQLTQEMMTEKERSNVVITRMKD RIGTLEKEHNVFQNKIHVSQYETQQMQMKFQQVREQMEAEIAHLKQENGILRDAVSNTTNQLE SKQSAELNKLQDYARLVNELTEKTGKLQQEEVQKKNAEQAAATQLKVQLQEAERRWEEVQSYI RKRTAEHEAAQQDLQSKFVAKENEVQSLHSLKLTDTLVSKQQLQRLMQLMESEQKRVNKEESL QMQRVDILEQNEALKAQIQQFHSQIAAQTSAVLAELHKKVIAEKDKQIKQTEDSLASERDRL TSKEEELKDIQNMNFKLAEVQKLQALANEQAAAAHELEKMQQSVYVKKDKIRLLEEQLQHEI SNKMEEFKILNDQNKALKSEVQKLQTLVSEQPNKDVVEQMEKCIQEKDEKLKTVEELLETLGLI QVATKEEELNARTENSSLTKEVQDLKAKQNDQVSFASLVEELKKVIEHKDGKIKSVEEELLEA ELLKVANKEKTVDLQKEIKALKEEIGNVQLEKAQQLSITSKVQELQNLKKGKEEQMNTMKAV LEEKEKDLANTGKWLQDLQEEENESLKAHVQEVQAQHNLEASSASQFEELEIVLKEKENELKRL EAMLERESDLSSKTQLLDVQDENKLFKSQIEQLKQNYQQASSFPPEELLKVISEREKEI SGLWNELDSLKDAVEHQKKNNERQQQVEAVELEAKEVLKKLFPKVSVPNSLSYGEWLHGFEK KAKECMAGTSGSEEVKVLEHKLKEADEMHLLQLECEKYKSVLAETEGILQKLQRSVEEENK WKVKVDESHKTIKMQSSFTSSEQELERLRSENKDIEHLRREREHLEMELEKAEMERSTYVTE VRELKAQLNETLTKLRTEQNERQKVAGDLHKAQSSLELIQSKIVKAAGDTTVIENSVDSPETE SSEKETMSVSLNQTVTQLQQLQAVNQQLTKEKEHYQVLE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 18B.

Table 18B. Comparison of NOV18a against NOV18b and NOV18c.		
Protein Sequence	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV18b	1..1328 1..1270	1041/1329 (78%) 1044/1329 (78%)
NOV18c	1..1328 1..1300	1062/1329 (79%) 1063/1329 (79%)

5

Further analysis of the NOV18a protein yielded the following properties shown in Table 18C.

Table 18C. Protein Sequence Properties NOV18a	
PSort analysis:	0.8200 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in plasma membrane; 0.1800 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 40 and 41

10

A search of the NOV18a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 18D.

Table 18D. Geneseq Results for NOV18a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB57163	Mouse ischaemic condition related protein sequence SEQ ID NO:396 - Mus musculus, 1327 aa. [WO200188188-A2, 22-NOV-2001]	1..1320 1..1326	1072/1354 (79%) 1183/1354 (87%)	0.0
AAG67538	Amino acid sequence of a human p180 protein - Homo sapiens, 1240 aa. [WO200164947-A1, 07-SEP-2001]	50..1068 235..1229	254/1069 (23%) 483/1069 (44%)	1e-63
AAM79523	Human protein SEQ ID NO 3169 - Homo sapiens, 1003 aa. [WO200157190-A2, 09-AUG-2001]	21..1068 51..992	261/1084 (24%) 482/1084 (44%)	3e-62
AAM78539	Human protein SEQ ID NO 1201 - Homo sapiens, 977 aa. [WO200157190-A2, 09-AUG-2001]	21..1068 25..966	261/1084 (24%) 482/1084 (44%)	3e-62
AAW89721	Canine ribosome receptor - Canis familiaris, 1484 aa. [WO9901565-A1, 14-JAN-1999]	36..982 511..1459	244/1006 (24%) 476/1006 (47%)	4e-62

In a BLAST search of public sequence databases, the NOV18a protein was found to have homology to the proteins shown in the BLASTP data in Table 18E.

5

Table 18E. Public BLASTP Results for NOV18a

Protein Accession Number	Protein/Organism/Length	NOV18a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q14707	156 kDa protein - Homo sapiens (Human), 1356 aa.	1..1328 1..1356	1328/1356 (97%) 1328/1356 (97%)	0.0
Q13999	CG1 protein (KIAA0004 protein) - Homo sapiens (Human), 1300 aa.	1..1328 1..1300	1297/1329 (97%) 1298/1329 (97%)	0.0
O97961	Kinectin - Vulpes vulpes (Red fox), 1330 aa.	1..1328 1..1330	1240/1330 (93%) 1284/1330 (96%)	0.0
Q61595	Kinectin - Mus musculus (Mouse), 1327 aa.	1..1320 1..1326	1072/1354 (79%) 1183/1354 (87%)	0.0
Q90631	Kinectin - Gallus gallus (Chicken), 1364 aa.	1..1328 1..1364	889/1370 (64%) 1119/1370 (80%)	0.0

Pfam analysis predicts that the NOV18a protein contains the domains shown in the Table 18F.

Table 18F. Domain Analysis of NOV18a			
Pfam Domain	NOV18a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LBP_BPI_CETP	576..608	8/35 (23%) 25/35 (71%)	0.27

5 Example 19.

The NOV19 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 19A.

Table 19A. NOV19 Sequence Analysis			
	SEQ ID NO: 67	921 bp	
NOV19a, CG11473-01 DNA Sequence	CAGACGCCTCCAGGATCTGTCGGCAGCTGCTGTTCTGAGGGAGAGCAGAGACCATGTCTGACA TAGAAGAGGTGGTGGAGAGTACGAGGAGGAGGAGCAGGAAGAAGCAGCTGTTGAAGAGCAGG AGGAGGCAGCGGAAGAGGATGCTGAAGCAGAGGCTGAGACCGAGGAGACCAGGGCAGAAGAAG ATGAAGAAGAAGAGGAAGCAAAGGAGGCTGAAGATGGCCCAATGGAGGAGTCCAAACCAAAGC CCAGGTCGTTTCATGCCCAACTTGGTGCCTCCCAAGATCCCGATGGAGAGAGAGTGGACTTTG ATGACATCCACCAGGAAGCGCATGGAGAAGGACCTGAATGAGTTGCAGGCGCTGATTGAGGCTC ACTTTGAGAACAGGAAGAAGAGGAGGAGGAGCTCGTTTCTCTCAAAGACAGGATCGAGAGAC GTCGGGCAGAGCGGGCCGAGCAGCAGCGCATCCGGAATGAGCGGGAAAAGAGAAGAAGATTTC TGGCTGAGAGGAGGAAGGTGCTGGCCATTGACCACCTGAATGAAGATCAGCTGAGGGAGAAGG CCAAGGAGCTGTGGCAGAGCATCTATAACTTGGAGGCAGAGAAGTTCGACCTGCAGGAGAAGT TCAAGCAGCAGAAATATGAGATCAATGTTCTCCGAAACAGGATCAACGATAACCAAGAAAGTCT CCAAGACCCGCGGAAGGCTAAAGTCACCGGGCGCTGGAATAGAGCCTGGCCTCCTTCACCA AAGATCTGCTCCTCGCTCGCACCTGCCTCCGGCCTGCACTCCCCCAGTTCCTGGGGCCTCCTG GGCACCCAGGCAGCTCCTGTTTGGAAATGGGGAGCTGGCCTAGGTGGGAGCCACCACTCCTG CCTGCCCCACACCACTCCACACCACTAATAAAAAGCC		
	ORF Start: ATG at 54		ORF Stop: TAG at 735
	SEQ ID NO: 68	227 aa	MW at 27175.7kD
NOV19a, CG11473-01 Protein Sequence	MSDIEEVVEEYEEEEQEEAAVEEQEEAAEEDAEAEAEETEETRAEEDEEEEEAKEEDGPMEEES KPKPRSFMPNLVPPKIPDGERVDFDDIHRKRMEKDLNELQALIEAHFENRKKEEELVSLKDR IERRRAERAEQQIRNREKKKKILAEERRKVLAIIDLNEDQLREKAKELWQSIYNLEAEKFDL QEFKQQKYEINVLNRINDNQVSKTRGKAKVTGRWK		
	SEQ ID NO: 69	1154 bp	
NOV19b, CG11473-02 DNA Sequence	CGGCCGCGTCGACAGCAGACGCTCCAGGATCTGTCGGCAGCTGCTGTTCTGAGGGAGAGCAG AGACCATGTCTGACATAGAAGAGGTGGTGGAGAGTACGAGGAGGAGGAGCAGGAAGAAGCAG CTGTTGAAGAGCAGGAGGAGGAGCAGCGGAAGAGGATGCTGAAGCAGAGGCTGAGACCGAGGAGA CCAGGGCAGAAGAAGATGAAGAAGAAGGAAGCAAAGGAGGCTGAAGATGGCCCAATGGAGG AGTCCAAACCAAAGCCAGGTCGTTTCATGCCCAACTTGGTGCCTCCCAAGATCCCGATGGAG AGAGAGTGGACTTTGATGACATCCACCAGGAAGCGCATGGAGAAGGACCTGAATGAGTTGCAGG CGCTGATTGAGGCTCACTTTGAGAACAGGAAGAAAGAGGAGGAGGAGCTCGTTTCTCTCAAAG ACAGGATCGAGAGACGTCGGGCAGAGCGGGCCGAGCAGCAGCGCATCCGGAATGAGCGGGAGA AGGAGCGGCAGAACCGCTGGCTGAAGAGAGGCTCGACGAGAGGAGGAGAGAGAGGAGGAGGAG AGGCTGAGGATGAGGCCCGGAAGAAGAAGGCTTTGTCCAACATGATGCATTTTGGGGTTACA TCCAGAAGCAGGCCACAGAGCGGAAAAGTGGGAAGAGGCAGACTGAGCGGAAAAGAAGA AGAAGATTCTGGCTGAGAGGAGGAAGGTGCTGGCCATTGACCACCTGAATGAAGATCAGCTGA GGGAGAAGGCCAAGGAGCTGTGGCAGAGCATCTATAACTTGGAGGCAGAGAAGTTCGACCTGC		

	AGGAGAAGTTCAAGCAGCAGAAATATGAGATCAATGTTCTCCGAAACAGGATCAACGATAACC AGAAAGTCTCCAAGACCCGCGGAAGGCTAAAGTCACCGGGCGCTGGAAATAGAGCCTGGCCT CCTTCACCAAGATCTGCTCCTCGCTCGCACCTGCCTCCGGCCTGCACTCCCCAGTTCCCGG GCCCTCCTGGGCACCCAGGCAGCTCCTGTTTGAAATGGGGAGCTGGCCTAGGTGGGAGCCA CCACTCCTGCCTGCCCCACACCCACTCCACACCAGTAATAAAAAGCCACCACACAAAAAA AAAAAAAAAAAAACCCAAA		
	ORF Start: ATG at 69		ORF Stop: TAG at 933
	SEQ ID NO: 70	288 aa	MW at 34589.9kD
NOV19b, CG111473-02 Protein Sequence	MSDIEEVVEEYEEEEQEEAAVEEQEEAAEEDAEAEAEETEETRAEEDEEEEAKEAEDGPMES KPKPRSFMPNLVPPKI PDGERVDFDDIHRKRMEKDLNELQALIEAHFENRKKEEELVSLKDR IERRRAERAEQQRIRNEREKERQNRLEAERARREEENRRKADEARKKKALSNMMHFGGYIQ KQAQTERKSGKRQTEREKKKILAEERRKVLADHLNEDQLREKAKELWQSIYNLEAEKFDLQE KFKQQKYEINVLRNRINDNQVSKTRGKAKVTGRWK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 19B.

Table 19B. Comparison of NOV19a against NOV19b.		
Protein Sequence	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV19b	128..227 190..288	88/100 (88%) 93/100 (93%)

5

Further analysis of the NOV19a protein yielded the following properties shown in Table 19C.

Table 19C. Protein Sequence Properties NOV19a	
PSort analysis:	0.9725 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV19a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 19D.

Table 19D. Geneseq Results for NOV19a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y91088	Recombinant modified human cardiac troponin T SEQ ID NO:6 - Homo sapiens, 288 aa. [US6060278-A, 09-MAY-2000]	1..227 1..288	221/288 (76%) 226/288 (77%)	e-113
AAB12186	Human troponin T cardiac isoform (cTnT) - Homo sapiens, 288 aa. [US6072040-A, 06-JUN-2000]	1..227 1..288	221/288 (76%) 226/288 (77%)	e-113
AA W72759	Recombinant human cardiac troponin T - Homo sapiens, 288 aa. [US5834210-A, 10-NOV-1998]	1..227 1..288	221/288 (76%) 226/288 (77%)	e-113
AA W41574	Human cardiac troponin T isoform T3 - Homo sapiens, 288 aa. [WO9739132-A1, 23-OCT-1997]	1..227 1..288	221/288 (76%) 226/288 (77%)	e-113
AA W76640	Human cardiac HcTnT protein deletion mutant delta S275-K288 - Homo sapiens, 274 aa. [DE19815128-A1, 08-OCT-1998]	1..213 1..274	205/274 (74%) 212/274 (76%)	e-104

In a BLAST search of public sequence databases, the NOV19a protein was found to have homology to the proteins shown in the BLASTP data in Table 19E.

Table 19E. Public BLASTP Results for NOV19a				
Protein Accession Number	Protein/Organism/Length	NOV19a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BUF6	Similar to troponin T2, cardiac - Homo sapiens (Human), 285 aa.	1..227 1..285	224/285 (78%) 225/285 (78%)	e-114
TPHUTC	troponin T, cardiac muscle - human, 298 aa.	1..227 1..298	221/298 (74%) 226/298 (75%)	e-110
AAK92232	Truncated cardiac troponin T - Homo sapiens (Human), 274 aa.	1..213 1..274	205/274 (74%) 212/274 (76%)	e-103
A25345	troponin T, cardiac muscle, major isoform - rabbit, 276 aa.	2..227 1..276	202/278 (72%) 212/278 (75%)	e-103
B25345	troponin T, cardiac muscle, minor isoform - rabbit, 276 aa.	2..227 1..276	197/278 (70%) 206/278 (73%)	3e-99

PFam analysis predicts that the NOV19a protein contains the domains shown in the Table 19F.

Table 19F. Domain Analysis of NOV19a			
Pfam Domain	NOV19a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Troponin	93..227	36/194 (19%) 105/194 (54%)	0.0067

5 Example 20.

The NOV20 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 20A.

Table 20A. NOV20 Sequence Analysis			
	SEQ ID NO: 71	5382 bp	
NOV20a, CG111501-01 DNA Sequence	AGGCTCCCAATCCCATCCTCATCTCTGCCCTTCTTCTCAGAAGGATGGCCGACACCCAGACA CAGGTGGCCCCACCAACCATGAGGATGGCAACTGCAGAGGACCTGCCCTCCCTCCACCC CCAGCCCTGGAGGACCTGCCACTGCCGCCACCAAGGAATCCTTCTCCAAGTCCATCAGCAG CGCAAGCTAGTGAGCTCCGCCCTCTACAGGCACATCCACCTGAGCTCCGCAAGAATCTG GCTGAGGCTGTGGCCGAGGATCTGGCTGAGGTCCTGGGCTCTGAGGAACCCACCGAGGTGAC GTTCAAGTGCATGCGCTGGATCTTTGAGAACTGGAGACTGGATGCCATTGGAGAACACGAGAG CCAGCTGCCAAGGAGCCCGTGTGTGGTGACGTCCAGGCCACCTCCCGCAAGTTTGAGGAA GGCTCCTTTGCCAACAGCACAGACCAGGAGCCAACAGGCCCCAGCCAGGTGGAGGAGACGTT CGTGACGCCCGCTGGCTATTGAGACAAAGCCACTGGACGAGCTGACAGGGCAAGCCAAGGAA CTGGAGGCCACTGTGAGGGAGCCTGCAGCCAGCGGAGATGTGAGGGTACCAGGATGCTCTTT GAGACGCGCCCGCTGGACCGCTGGGCTCCCGCCCTCCCTGCAGGACAGAGCCCTTGGAA CTGCGCTCAGAGATCCAGGAGCTGAAGGGTGATGTGAAAAAGACAGTGAAGCTCTTCCAAACG GAGCCCTGTGTGCCATCCAGGATGCAGAGGGCGCCATCCATGAGGTCAAGGCCGATGCCGG GAGGAGATCCAAAGCAACGCGGTGAGGTCTGCCCGCTGGCTCTTTGAGACCCGGCCTCTGGAC GCCATCAACCAGGACCCAGCCAGGTGCGGGTGATCCGGGGGATTTCCTGGAGGAGGGGGCC CGGCCCCGACGTGAGTGCAACTCGCTGGATCTTTGAGACACAGCCCTGGATGCCATCCGGGAG ATCTTGGTAGATGAGAAGGACTTCCAGCCATCCCCAGACCTTATCCACCTGGTCCAGATGTT CAGCAGCAGCAGCATCTGTTGAGACCCGAGCGCTGGACACTCTGAAGGGGACGAAGAGGCT GGAGCAGAGGCCCCACCAAGGAGGAAGTGGTCCCTGGTGATGTCCGCTCCACCCTGTGGCTA TTTGAACAAAGCCCTGGATGCTTTCAGAGACAAGGTCCAAGTGGGTCACTACAGCGAGTG GATCCCCAGGACGGTGAGGGGCATCTATCCAGTGACAGCTCCTCAGCACTGCCCTTCTCTCAG AGTGCCCCCAGAGGGATGAGCTAAAGGGGGATGTGAAGACTTTTAAGAACCTTTTGTAGACC CTTCCCTTGGACAGCATTGGACAGGGTGAGGTTCTGGCCCATGGGAGTCCAAGCAGAGAAGAA GGAAGTATTCTGTGGGAGGCCAGGGCATAGGGTCCCAAGTGTATGCCATCAGGACAGC AAGGGCCGCTCCATGCCCTGACCTCTGTAGCAGAGAGCAGATAGTCGGAGGTGATGTGCAG GGCTACAGGTGGATGTTTGAACACAGCCCTAGACCAGCTCGGCCGAAGCCCCAGTACCATC GACGTGTGCGGGGCATACCCGCGCAGGAAGTGGTGGCTGGGGACGTTGGCACAGCTCGGTGG CTTTTGAACCCAGCCCTGGAGATGATCCACCAACGGGAGCAGCAGGAACGACAGAAAGAA GAAGGAAGAGTCAGGGAGACCCAGCCCTGAGGCACCCCAAAGGGCGATGTGCAGACCATC CGGTGGTGTTCGAGACTTGCCCAATGAGTGAGTTGGCCGAAAAGCAGGGGTGAGAGTCA GATCCCAAGCCAAAGGCTGAGGCACAGTCTGCACCTGGATGTTCAAGCCCCAACCTGTGGAC AGGCCAGTGGGCTCCAGGGAGCAGCACCTGCAGGTAGCCAGGTCCCGGCTGGGGAAGACAG ACAGACAGACACGTCTTTGAGACCGAGCCTTTCAGGCCCTCAGGCCGTCCCTGTGGAAGACGG CCTGTGAGATACTGCAGCCCGCTGGAGATCCCTTCAGGGCAGGTGTCTCGTCAGAAAGAGTT TTTCAGGCCCTGGAGGCAGGCAAGAAGGAAGAACAGGAGCCCCGGTAATCGCTGGGTCCATC CCCGCGGTTCTGTCCACAAGTTCACTTGGCTTTTGAAGATTGTCCCATGGGCTCCCTGGCA GCTGAGAGCATCCAAGGGGCAACCTCCTGGAAGAGCAGCCATGAGCCCTCAGGCAACAGG		

	ATGCAAGAGAGCCAGGAGACTGCAGCTGAGGGGACCCTGCGGACTCTGCATGCCACACCTGGC ATCTTGCACCATGGAGGCATCTCATGGAGGCCGAGGGCCAGGGGAGCTCTGTCTTGCCAAG TATGTGCTCTCGGGCACAGGGCAGGGGCACCCCTTATATACGAAAGGAGGAGCTGGTGTGAGGT GAACTTCCAGGATCATCTGCCAAGTCTTGCCTGCGCCGCCAGATGTGGACCAAGGGGCTGTCTG GTGCAGGAAGACCACTGGCCAGCTCCAACCTCAAGCCGCTGAGGCTGCCAAGTCCAGGCAGC AGTGGGAATATTGAAGACATGGACCCTGAGCTCCAGCAGCTGCTGGCTTGGGCTTGGGACC TCCGTGGCAAGGACTGGGCTGGTGATGCAGGAGACAGAGCAGGGCCTGGTGCAGCTGACTGCC TACTCTCTGCAGCCCCGGCTAACTAGCAAGGCCTCTGAGAGGAGCAGCGTGCAGCTGTTGGCC AGCTGCATAGATAAAGGAGACCTGAGTGGCTGCACAGTCTGCGGTGGGAGCCCCGGCTGAC CCGAGTCCAGTGGCAGCCAGCGAGGGGGCCAGAGCCTGCACCAACTGAGAGCATCATCCAT GTTCCCCACTGGACCCAGCATGGGGATGGGGCATCTGAGAGCCTCAGGGGCCACCCCTTGC CCTCTCAGGCCATTGGAAGGCAGTCCCTCTGGCTGGGGAGCTGCAGCACCAGCCCAATTG CAAAACACAGAAAAGCAGGAAGACAGTCACTCTGGACAGAAAGGGATGGCAGTCTTGGGAAAG TCAGAAGGAGCCAGCACTACCCCTCCGGGGCCTGGGGCCCCAGACCTCTGGCCGCCATGCAG AGTCTGCGGATGGCAACAGCTGAAGCCCAGAGCCTGCACCAGCAAGTCTGAACAAGCACAAAG CAGGGCCCCACCCCAACGCCACTTCCAACCCCATCCAGGACGGTCTTCGGAAGCTGGGGCT ACCCAAAGCAACATAAGGCTGGGGGTGGAAGTGATCCCCGGATCCAGCAGCCCCCAGAAAG GTCACTCTGACTTTCAGCTGGAGCCACCGTGTGAGGACTCCATCCAGCAAGCCTCTGAG CCCCTGAAGGACCCCTTCTTCACTCCACAGCAGCCCTGCTGGCCAGAGAACCCTGGAGGG TCACAGACAAAGACCCCAAACTGGACCCACCATGCCCCAAAGAAGAAGCCGAGCTGCCC CCTAAACCTGCACACCTAACCCAGAGCCACCCCTCTCAGAGGCTGCCAAGCCCTTGCTCTA TCTCCAGCTTTTCTCGAGGTGGGGCAAAGAGAACACCAACGAGGTGAGAGAGATACAGCC ATCCCTCAGCCAGCCAAAGTTCCCACTACTGTAGACCAGGGCCACATACCTCTGGCCAGATGT CCCAGTGGACATAGCCAGCCAGCTTACAACATGGCCTCAGCACCACGGCCCCCAGGGCCACC AAGAATCAGGCTACAGGCAGCAATGCCAGAGCTCTGAGCCCCCAAGCTCAATGCCCTCAAC CATGATCCACCTCACCACAGTGGGGCCCCGGCCCTCAGGAGAGCAGCCCATGGAAGGTTCC CACCAAGGGGCCCTGAGAGCCCTGACAGTCTGCAAGAAGAACAGAAAGAGTCCAGGCTCCAG CTGAACCAGGTGCAAGCCCTGGAGAAGGAGGCCGCAAGCAGTGTGGACGTGACGGCCCTGCGG AGGCTCTTTGAGGCCGTGCCCCAGCTGGGAGGGGTGCTCCTCAGGCTCTGCTGCCACCAA AAGCCCCAGGCCCTCAGTGGAGCAGGCCCTTGGGGAGCTGACACGGGTGAGCAGGAAAGTTGCT CAACTGAAGGAACAGACCTTGGCAAGGCTGCTGGACATTGAAGAGGCTGTGCACAAGGCACCT AGCTCCATGTCTAGCCTCCAGCCTGAGGCCAGTGCACAGAGCCATTTCAGGGACCTCCAAAA GACCACAGTGGCCACAAGATCAGTGTACAGTACAGTACAGCAGGCCCCAGTGGCTCAGGC CAGGAGGTGCGAGGTCAAACCTGAGTCAAGAACCAAGCCAGGTTGAATGCCACACTGAGGCC CAGAGTCAAGTCAAGATCAGAAATCACACAGAGGCCAGAGGTACACAGCCTCAACTGCCCTT TCCACCAGGAGGCAGGAGACATCAAGAGAGTATTGTGCTCCTCGGGTTTCACTTCCAGC CGAGATTCTCCCTCCTCCCAACATTTATCTCCATCCAGTCCGCCACAAGGAAGCCTCTAGAG ACTCCAGCTTTAAGGGCAACCTGATGTCTCAGTGAAGACACACAAGTGGCTCAGGACATA GGCCAGGCCCTGCTCCACAGAAAGGTGTCCAAGACAAAAGTGGGAAGAGGACATACCCAG TGCTCTGTGCAACCTGAACCTGCCCCCTCCCTCAGCCAGTCCCCTGCCAGAGGTGGCAAAAG AGTGTCTGGAGCTACAGACGGGGCCAGGGAGCTCAACAACATATGGAGCATGAGAACCGTG ACTGAACAGTATGAGGAGGTGGACAGTTTGGGAACACAGTCTCATGTCTTCCACCACAGTC ACCGAGCAGGCAGAGCCACCCAGGAACCCAGGCTCCACCTCGGGCTCCAGCCTCCCCCTTG CTGAGGCAGTTCTGCACAGCCAGCTGGGTTCAGCAGTGACCTGACAGAAGCTGAGACGGTG CAGGTGTCTGCAGTACTCCAGCCAGCTGCCAGTGAAGGCCACCCCTCCACACACCT GCCACCTGTTCTGGCTCCACTGCCCCAGGACTGAAGTGGGTACCTGCCTCCTGTACACTGG AGCAAGGACCAAGAGGAAATGGCATCTTCAGAGGATTACTGTGGGCCATTTCCTTTTCGAGT TCTTTCAATAGGCCAGTTCTTCCAAATGGAAAAAGAAAGGTCTGGAAGAGGCCACAGAGTT GCACAGGCGTGGGGTAGGATGGGGC		
	ORF Start: ATG at 46		ORF Stop: TGA at 5140
	SEQ ID NO: 72	1698 aa	MW at 183686.6kD
NOV20a, CG11501-01 Protein Sequence	MADTQTVAPTPTMRMATAEDLPLPPPALEDLPLPPKESFSKFHQQRQASELRRLRYRHIHP ELRKNLAEEVAEDLAEVLGSEETPEGDVQCMRWIFENWRLDAIGEHERPAAKEPVLCDGVQAT SRKFEESGFANSTDQEPTRPQGGDVRAARWLFETKPLDELTGQAKELEATVREPAASGDVQ GTRMLFETRPLDRLGSRPSLQESPLRLSEIQLKGDVKKTVKLFQTEPLCAIQDAEGAIHE VKAACREEIQSNVARSARWLFETRPLDAINQDPSQVRVIRGISLEEGARPDSATRWIFETQP LDAIREILVDEKDFQPSPLIPPGPDVQQQHLFETRALDTLKGDEEAGAEAPPKEEVVPGDV RSTLWLFETKPLDAFRDKVQVGHLLQVDPQDGEHLSSDSSSALPFSQSAPQRDELKGDVKT FKNLFETLPLDSIGQEVLAHGSFSEEGTDSAGQAQIGSPVYAMQDSKGRHLTSSVREQI VGGDVQGYRWMFETQPLDQLGRSPSTIDVVRGITRQEVVAGDVGTARWLFETQPLEMIHQREQ QERQKEEGKSQGDPPQEPKGDVQTIKWLFTCPMSLAELKQGEVTDPTAKAEASQCTWMP KPQPVDRPVGSREQLQVSQVPAGERQTDHVFETEPQLQASGRPCGRRPVRYCSRVEIPSGQV SRQKEVFQALEAGKKEEQPRVIAGSIPAGSVHKFTWLFENCMPMSLAELKQGEVTDPTAKAEASQCTWMP SPNGNRMQESQETAAGTTLRLHATPGILHHGILMEARGPGELCLAKVVLSTGGHLYPIRK EELVSGELPRIICQVLRPPVDVQQLLVQEDPTGQLQLKPLRLPTPGSSGNIEDMDPELQQLL		

	ACGLGTSVARTGLVMQETEOGLVALTAYSLOPRLTSKASERSVQLLASCIDKGDLSGLHSLR WEPPADPSPVPASEGAQSLHPTESIIHVPPLDPSMGMHLRASGATPCPPQAIGKAVPLAGEA AAPAQLQNTKEQEDSHSGQKMAVLGKSEGATTTTPGPGAPDLLAAMQSLRMATAEAQSLHQQ VLNKKHQGPTPTATSNPIQDGLRKAGATQSNIRPGGSDPRIAAPRKVSPDFPAGAHRAEDS IQQASEPLKDPPLLHSHSSPAGQRTPGGSQTKTPKLDPTMPKKPKLPKPAHLTQSHPPQRL PKPLPLSPSFSSEVQGQREHQRGERDTAIPQPAKVPTTVQDGHIPARCPSGHSQPSLQHGLST TAPRPTKNQATGSNAQSSEPPKLNALNHDPTSPQWGPSPGSEQPMESHQGAPESPDSLQRNQ KELQGLLNQVQALEKEAASSVDVQALRRLFEAVPQLGGAAPQAPAAHQPEASVEQAFGELTR VSTEVAQLKEQTLARLLDIEEAVHKALSSMSSLOPEASARGHFQGPDKHSAHKISVTVSSSA RPSGSGQEVGGQTAVKNOAKVECHTEAQSOVKIRNHTAARGHTASTAPSTRRQETSREYLCP RVLPSSRDSPPSPTFISIQSATRKPLETPSFKGNPDVSVKSTQLAQDIGQALLHQKGVQDKT KKDITQCSVQPEPAPPSASPLPRGWQKSVLELQTGPGSSQHYGAMRTVTEQYEEVDQFGNTVL MSSTTVTEQAEPNRNPGSHLGLHASPLLRQFLHSPAGFSSDLTEAETVQVSCSYSQPAAQ		
	SEQ ID NO: 73	3333 bp	
NOV20b, CG111501-02 DNA Sequence	AGGCTCCCAATCCCATCCTCATCTCTGCCCCCTTCTTCTCAGAAGGATGGCCGACACCCAGACA CAGGTGGCCCCACCAACCATGAGGATGGCAACTGCAGAGGACCTGCCCTCCCTCCACCC CCAGCCCTGGAGGACTGCCACTGCCGACCCACCAAGGAATCCTTCTCCAAAGTCCCATCAGCAG CGGCAAGCTAGTGAGCTCCGCCGCTCTACAGGCACATCCACCTGAGCTCCGCAAGAACTCTG GCTGAGGCTGTGGCCGAGGATCTGGCTGAGGTCTCTGGGCTCTGAGGAACCCACCGAGGGTGAC GTTCAAGTGCATGCGCTGGATCTTTGAGAACTGGAGACTGGATGCCATTGGAGAACACGAGAGG CCAGCTGCCAAGGAGCCCGTGCTGTGTGGTGACGTCCAGGCCACCTCCGCAAGTTTGAGGAA GGCTCCTTTGCCAACAGCACAGACCAGGAGCAACCAAGGCCAGCCAGGAGGAGAGACGTT CGTGACGCCGCTGGCTATTTGAGACAAAGCCACTGGACGAGCTGACAGGGCAAGCCAAGGAA CTGAGGGCCACTGTGAGGGAGCCTGCAGCCAGCGGAGATGTGAGGGTACCAGGATGCTCTTT GAGACGCGGCCGCTGGACCGCTGGGCTCCCGCCCCCTCCCTGCAGGAGCAGAGCCCCCTTGAA CTGCGCTCAGAGATCCAGGAGCTGAAGGGTGATGTGAAAAGACAGTGAAGCTCTTCCAAACG GAGCCCCCTGTGTCCATCCAGGATGCAGAGGGCGCCATCCATGAGGTCAAGGCCGATGCCGG GAGGAGATCCAAAGCAACGCGGTGAGGTCTGCCCGCTGGCTCTTTGAGACCCGGCCCTCTGGAC GCCATCAACCAGGACCCCAAGCCAGGTGCGGGTGATCCGGGGGATTTCCCTGGAGGAGGGGGCC CGGCCGACGTCAGTGCAACTCGCTGGATCTTTGAGACACAGCCCCCTGGATGCCATCCGGGAG ATCTTGGTAGATGAGAAGGACTTCCAGCCATCCCCAGACCTTATCCCACTGGTCCAGATGTT CAGCAGCAGCAGCATCTGTTTGAACCCGAGCGCTGGACACTCTGAAGGGGGACGAAGAGGCT GGAGCAGAGGCCCCACCAAGGAGGAAGTGGTCCCTGGTGATGTCCGCTCCACCCTCTGGGCTA TTTGAAACAAAGCCCCCTGGATGCTTTGAGAGACAAGGTCCAAGTGGGTACCTACAGCGAGTG GATCCCCAGGACGGTGAGGGGCATCTATCCAGTGACAGCTCCTCAGCACTGCCCTTCTCTCAG AGTGCCCCCAGAGGGATGAGCTAAAGGGGGATGTGAAGACTTTTAAGAACCTTTTGAAGACC CTTCCCTTGGACAGCATTGGACAGGGTGAGGTTCTGGCCCATGGGAGTCCAAGCAGAGAAGAA GGAACTGATTCTGCTGGGCGAGGCCAGGGCATAGGGTCCCCAGTGTATGCCATGCAGGACAGC AAGGGCCGCTCCATGCCCTGACCTCTGTTAGCAGAGAGCAGATAGTCGAGGTGATGTGCAG GGCTACAGGTGGATGTTTGAACACAGCCCTAGACCAGCTCGGCCGAAGCCCCAGTACCATC GACGTGGTGCGGGGCATCACCCGGCAGGAAGTGGTGGCTGGGGACGTTGGCACAGCTCGGTGG CTTTTGAAGACCCAGCCCCCTGGAGATGATCCACCAACGGGAGCAGCAGGAACGACAGAAAGAA GAAGGGAAGAGTCAGGAGACCCCCAGCCTGAGGCACCCCCAAAGGGCGATGTGCAGACCATC CGGTGGTTGTTTCGAGACTTGCCCAATGAGTGAGTTGGCCGAAAAGCAGGGGTGAGAGGTACA GATCCCAAGCCAAAGGCTGAGGCACAGTCTGCACCTGGATGTTCAAGCCCCAACCTGTGGAC AGGCCAGTGGGCTCCAGGGAGCAGCACCTGCAGGTAGCCAGGTCCCGGCTGGGGAAGACAG ACAGACAGACAGCTTTTGAACCGAGCCTCTTCAAGGCTCAGGCCGCTCCCTGTGGAAGACGG CCTGTGAGATACTGCAGCCGCGTGGAGATCCCTTCAAGGCAGGTGTCTGTCAGAAAGAGGTT TTTCAAGGCCCTGGAGGCAGGCAAGAAGGAAGAACAGGAGCCCCGGGTAATCGCTGGGTCCATC CCCGCGGGTCTGTCCACAAGTTCACTTGGCTTTTGAAGATTTGCCATGGGCTCCCTGGCA GCTGAGAGCATCAAGGGGGCAACCTCTGGAAGAGCAGCCATGAGCCCTCAGGCAACAGG ATGCAAGAGAGCCAGGAGACTGCAGCTGAGGGGACCCCTGCGGACTCTGCATGCCACACCTGGC ATCCTGCACCATGGAGGCATCCTCATGGAGGCCGAGGGCCAGGGGAGCTGTGCTTGGCAAG TATGTGCTCTCGGGCACAGGGCAGGGGACCCCTTATATACGAAAGGAGGAGCTGGTGTGAGGT GAACTTCCAGGATCATCTGCCAAGTCTTGCGCCGGCCAGATGTGGACACAGCGGGCTGTCT GTGCAGGAAGACCAACTGGCCAGCTCCAATCAAGCCGCTGAGGCTGCCAACTCCAGGCAGC AGTGGGAATATTGAAGACATGGACCTGAGCTCCAGCAGCTGCTGGCTTGGCGTCTTGGGACC TCCGTGGCAAGGACTGGGCTGGTGATGCAGGAGACAGAGCAGGGCCTGGTGCACACTGACTGCC TACTCTCTGCAGCCCCGGCTAACTAGCAAGGCCTCTGAGAGGAGCAGCGTGACGCTGTGGCC AGCTGCATAGATAAAGGAGACCTGAGTGGCCTGCACAGTCTGCGTGGGAGCCCCCGGCTGAC CCGAGTCCAGTGCCAGCCAGCGAGGGGGCCAGAGCCTGCACCAACTGAGAGCATCATCCAT GTTCCCCCACTGGACCCCAACAGCCACTTCAACCCCATCCAGGACGGTCTTCGGAAGCTGG GGCTACCCAAAGCAACATAAGGCCTGGGGGTGGAAGTGATCCCCGGATCCCAAGCAGCCCCAG AAAGCTCTGTGACAGGACTGACTTTCCAGCTGGAGCCACCGTCTGAGGACTCCATCCAG CAAGCCTCTGAGCCCCGTAAGGACCCCCCTTCTCACTCCACAGCAGCCCTGCTGGCCAGAGA		

	ACCCCTGGAGGGTCACAGACAAAGACCCCAAACCTGGACCCACCATGCCCCAAAGAAG CCGCAGCTGCCCCCTATATCTGCACACCTAACCCAGAGCCCCCTCCTCAGAGGCTG		
	ORF Start: ATG at 46		ORF Stop: TGA at 3061
	SEQ ID NO: 74	1005 aa	MW at 110888.2kD
NOV20b, CG111501-02 Protein Sequence	MADTQTQVAPTPTMRMATAEDLPLPPPALEDLPLPPPKESFSKFHQQRQASELRRLRYRHIHP ELRKNLAEVAEADLAEVLGSEEPTEGDVQCMRWIFENWRLDAIGEHERPAAKEPVLCDVQAT SRKFEEGFSFANSTDQEPTRPQPGGVDVRAARWLFETKPLDELGTQAKELEATVREPAASGDVQ GTRMLFETRPLDRLGSRPSLQEQSPLELRSEIQELKGDVKKTVKLFQTEPLCAIQDAEGAIE VKAACREEIQSNAVSARWLFETRPLDAINQDPSQVRVIRGISLEEGARPDVSATRWIFETQP LDAIREILVDEKDFQSPDLIPPGPDVQQQHLFETRALDTLKGDEEAGAEAPPKEEVVPGDV RSTLWLFETKPLDAFRDKVQVGHLLQRVDPQDGEHGLSSDSSSALPFSQAPQRDELKGDVKT KNLFETLPLDSIGQGEVLAHGSPSREEGTDGSAQAQIGSPVYAMQDSKGRHLALTSSVREQI VGVDVQGYRWMFETQPLDQLGRSPSTIDVVRGITRQEVVAGDVGTARWLFETQPLEMIHQREQ QERQKEEGKSQGDQPPEAPKGDVQTIKRLWLFETCPMSELAEKQSEVTDPTAKAEAQSTWFM KPQPVDRPVGSREQHLQVSQVPAGERQTDHVFETEPLQASGRPCGRPRVYCSRVEIPSGQV SRQKEVFQALEAGKKEEQEPRVIAGSIPAGSVHKFTWLFENCPMGSAAESIQQGNLLEEQPM SPSGNRMQESQETAEGTLRLTHATPGILHHGGILMEARGPGELCLAKYVLSGTGQGHPIYRK EELVSGELPRIICQVLRPDVDQQLLVQEDPTGQLQLKPLRLPTPGSSGNIEDMDPELQQLL ACGLGTSVARTGLVMQETEQGLVALTAYSLQPRLTASKASERSVQLLASCIDKGDLSGLSLR WEPPADPSPVPASEGAQSLHPTESIIVPPLDPNSHFQPHGRSSSESWGYPKQHKAWGK		
	SEQ ID NO: 75		1819 bp
NOV20c, 249257832 DNA Sequence	CACCAAGCTTATGGCCGACACCCAGACACAGGTGGCCCCACACCAACCATGAGGATGGCAAC TGCAGAGGACCTGCCCTCCCTCCACCCAGCCCTGGAGGACCTGCCACTGCCGCCACCCAA GGAATCCTTCTCCAAGTTCATCAGCAGCGGCAAGCTAGTAGCTCCGCCGCCCTCTACAGGCA CATCCACCTGAGCTCCGCAAGAATCTGGCTGAGGCTGTGGCCGAGGATCTGGCTGAGGTCTCT GGGCTCTGAGGAACCCAGGAGGTGACGTTCACTGCATGCGCTGGATCTTTGAGAACTGGAG ACTGGATGCCATTGGAGAACACGAGAGGCCAGCTGCCAAGGAGCCCGTGTGTGTGGTGACGT CCAGGCCACCTCCCGCAAGTTTGAAGAAAGGCTCCTTTGCCAACAGCACAGACCAGGAGCCAAC CAGGCCCCAGCCAGGTGGAGGAGACGTTCTGTCAGCCCGCTGGCTATTGAGACAAAGCCACT GGACGAGCTGACAGGGCAAGCCAAGGAACCTGTGAGGAGCCTGCAGGAGCCTGCAGCAGCGG AGATGTGCAGGGTACCAGGATGCTCTTTGAGACGCGCGCTGGACCGCCTGGGCTCCCGCCC CTCCCTGCAGGAGCAGAGCCCTTGAACCTGCGCTCAGAGATCCAGGAGCTGAAGGGTGATGT GAAAAAGACAGTGAAGCTCTTCCAACCGAGCCCCCTGTGTGCCATCCAGGATGCAGAGGGCGC CATCCATGAGGTCAAGGCCGATCGCCGGAGGAGATCCAAAGCAACGCGGTGAGGTCTGCCCG CTGGCTCTTTGAGACCCCGGCTCTGGACGCCATCAACAGGACCCAGCCAGGTCGGGTGAT CCGGGGGATTTCCTGGAGGAGGGGGCCCGGCCGACGTCAGTGCAACTCGCTGGATCTTTGA GACACAGCCCTGGATGCCATCCGGGAGATCTTGGTAGATGAGAAGGACTTCCAGGCATCCCC AGACCTTATCCCACTGGTCCAGATGTTACAGCAGCAGCGGCATCTGTTGAGACCCGAGCGCT GGACACTCTGAAGGGGACGAAGAGGCTGGAGCAGAGGCCCAACCAAGGAGGAAGTGGTCCC TGGTGATGTCCGCTCCACCTGTGGCTATTTGAAACAAAGCCCTGGATGCTTTCAGAGACAA GGTCCAAGTGGGTACCTACAGCAGTGGATCCCCAGGACGGTGAGGGGCATCTATCCAGTGA CAGCTCCTCAGCACTGCCCTTCTCTCAGAGTGCCCCCAGAGGGATGAGCTAAAGGGGGATGT GAAGACTTTTAAGAACCTTTTGAAGACCTTCCCTTGGACAGCATTTGACAGGGTGAGGTTCT GGCCCATGGGAGTCCAAGCAGAGAGAAGGAAGTATTCTGCTGGGACAGGCCAGGGCATAGG GTCCCCAGTGATGCCATGCAGGACAGCAAGGGCCGCTCCATGCCCTGACCTCTGTAGCAG AGAGCAGATAGTCGAGGTGATGTGCAGGGCTACAGGTGGATGTTTGAGACACAGCCCTAGA CCAGCTCGGCCGAAGCCCCAGTACCATCGACGTGGTGCGGGGCATACCCGGCAGGAAGTGGT GGCTGGGGACGTTGGCACAGCTCGGTGGCTTTTGAAGACAGCCCTGGAGATGATCCACCA ACGGGAGCAGCAGGAACGACAGAAAGAAGAAGGAAGAGTCAGGGAGACCCCGAGCCTGAGGC ACCCCAAGGGCGATGTGCAGACCATCCGGTGGTGTTCGAGACTCTCAGGGC		
	ORF Start: at 2		ORF Stop: end of sequence
	SEQ ID NO: 76	606 aa	MW at 67469.6kD
NOV20c, 249257832 Protein Sequence	TKLMADTQTQVAPTPTMRMATAEDLPLPPPALEDLPLPPPKESFSKFHQQRQASELRRLRYRHI HPELRKNLAEVAEADLAEVLGSEEPTEGDVQCMRWIFENWRLDAIGEHERPAAKEPVLCDV QATSRKFEEGFSFANSTDQEPTRPQPGGVDVRAARWLFETKPLDELGTQAKELEATVREPAASG DVQGTMLFETRPLDRLGSRPSLQEQSPLELRSEIQELKGDVKKTVKLFQTEPLCAIQDAEGA IHEVKAACREEIQSNAVSARWLFETRPLDAINQDPSQVRVIRGISLEEGARPDVSATRWIFE TQPLDAIREILVDEKDFQSPDLIPPGPDVQQQHLFETRALDTLKGDEEAGAEAPPKEEVV GDVRSSTLWLFETKPLDAFRDKVQVGHLLQRVDPQDGEHGLSSDSSSALPFSQAPQRDELKGDV KTFKNLFETLPLDSIGQGEVLAHGSPSREEGTDGSAQAQIGSPVYAMQDSKGRHLALTSSVSR EQIVGGDVQGYRWMFETQPLDQLGRSPSTIDVVRGITRQEVVAGDVGTARWLFETQPLEMIHQ REQQERQKEEGKSQGDQPPEAPKGDVQTIKRLWLFETLEG		

	SEQ ID NO: 77	1216 bp
NOV20d, 249263153 DNA Sequence	CACCAAGCTTTCAGGAGAGCAGCCCATGGAAGGTTCCCAAGGGGCCCTGAGAGCCCTGA CAGTCTGCAAAGAAACCAGAAAGAGCTCCAGGGCCTCCTGAACCAGGTGCAAGCCCTGGAGAA GGAGGCCGCAAGCAGTGTGGACGTGCAGGCCCTGCGGAGGCTCTTTGAGGCCGTGCCCCAGCT GGGAGGGGCTGCTCCTCAGGCTCCTGCTGCCACCAAAGCCGAGGCTCAGTGGAGCAGGC CTTTGGGGAGCTGACACGGGTGACACGGAAGTTGCTCAACTGAAGGAACAGACCTTGGCAAG GCTGCTGGACATTGAAGAGGCTGTGCACAAGGCACTCAGCTCCATGTCTAGCCTCCAGCCTGA GGCCAGTGCCAGAGGCCATTTCAGGGACCTCCAAAAGACCAGTGCCCAAGATCAGTGT CACAGTCAGCAGTAGCGCCAGGCCAGTGGCTCAGGCCAGGAGGTGAGAGTCAAAGTGCAGT CAAGAACCAGCCAGGTTGAATGCCACACTGAGGCCAGAGTCAAGTCAAGATCAGAAATCA CACAGAGGCCAGAGGTACACAGCCTCAACTGCCCCTTCCACCAGGAGGCAGGAGACATCAAG AGAGTATTTGTGCCCTCCTCGGGTTTACCTTCCAGCCGAGATTCTCCCTCCTCCCAACATT TATCTCCATCCAGTCGGCCACAAGGAAGCCTCTAGAGACTCCAGCTTTAAGGGCAACCCTGA TGTCTCAGTGAAAAGCACACAAGTGGCTCAGGACATAGGCCAGGCCCTGCTCCACCAGAAAGG TGTCGAAGACAAAAGTGGGAAGAAGGACATCAGGAGTGTCTGTGCAACCTGAACCTGCCCC TCCTCAGCCAGTCCCTGCCCAGAGGGTGGCAAAAGAGTGTCTGGAGCTACAGACGGGGCC AGGGAGCTCACAACACTATGGAGCCATGAGAACCCTGACTGAACAGTATGAGGAGGTGGACCA GTTTGGGAACACAGTCTCATGTCTTCCACCAGTACCGAGCAGGCAGAGCCACCCAGGAA CCCAGGCTCCACCTCGGGTCCACGCCTCCCCCTTGCTGAGGCAGTTCTGTCAGAGCCAGC TGGGTTGAGCAGTGACCTGACAGAAGCTGAGACGGTGCAGGTGTCTGCAGTACTCCAGCC AGCTGCCAGCTCGAGGGC	
	ORF Start: at 2	ORF Stop: end of sequence
	SEQ ID NO: 78	405 aa MW at 43419.8kD
NOV20d, 249263153 Protein Sequence	TKLSGEQPMEGSHQGAPESPDLSQRNQKELQGLLNQVQALEKEAASSVDVQALRRLFEAVPQL GGAAPQAPAAHQKPEASVEQAFGELTRVSTVEVAQLKEQTLARLLDIEAVHKALSSMSLQPE ASARGHFQGPDKDHAHKISVTVSSSARPSGSGQEVGQTAVKNQAKVECHTEAQSQVKIRNH TEARGHTASTAPSTRRQETSREYLCPPRVLPSRDSPPSSPTFISIQSATRKPLETPSFKNPD VSVKSTQLAQDIGQALLHQKGVQDKTITQCSVQPEPAPPSASPLPRGWQKSVLELQGTGP GSSQHYGAMRTVTEQYEEVDQFGNTVLMSSTTVTEQAEPNPNPGSHLGLHASPLLRQFLHSPA GFSSDLTEAETVQVSCSYSQPAAQLEG	
	SEQ ID NO: 79	1216 bp
NOV20e, 249263166 DNA Sequence	CACCAAGCTTTCAGGAGAGCAGCCCATGGAAGGTTCCCAAGGGGCCCTGAGAGCCCTGA CAGTCTGCAAAGAAACCAGAAAGAGCTCCAGGGCCTCCTGAACCAGGTGCAAGCCCTGGAGAA GGAGGCCGCAAGCAGTGTGGACGTGCAGGCCCTGCGGAGGCTCTTTGAGGCCGTGCCCCAGCT GGGAGGGGCTGCTCCTCAGGCTCCTGCTGCCACCAAAGCCGAGGCTCAGTGGAGCAGGC CTTTGGGGAGCTGACACGGGTGACACGGAAGTTGCTCAACTGAAGGAACAGACCTTGGCAAG GCTGCTGGACATTGAAGAGGCTGTGCACAAGGCACTCAGTCCATGTCTAGCCTCCAGCCTGA GGCCAGTGCCAGAGGCCATTTCAGGGACCTCCAAAAGACCAGTGCCCAAGATCAGTGT CACAGTCAGCAGTAGCGCCAGGCCAGTGGCTCAGGCCAGGAGGTGAGAGTCAAAGTGCAGT CAAGAACCAGCCAGGTTGAATGCCACACTGAGGCCAGAGTCAAGTCAAGATCAGAAATCA CACAGAGGCCAGAGGTACACAGCCTCAACTGCCCCTTCCACCAGGAGGCAGGAGACATCAAG AGAGTATTTGTGCCCTCCTCGGGTTTACCTTCCAGCCGAGATTCTCCCTCCTCCCAACATT TATCTCCATCCAGTCGGCCACAAGGAAGCCTCTAGAGACTCCAGCTTTAAGGGCAACCCTGA TGTCTCAGTGAAAAGCACACAAGTGGCTCAGGACATAGGCCAGGCCCTGCTCCACCAGAAAGG TGTCGAAGACAAAAGTGGGAAGAAGGACATCAGGAGTGTCTGTGCAACCTGAACCTGCCCC TCCTCAGCCAGTCCCTGCCCAGAGGGTGGCAAAAGAGTGTCTGGAGCTACAGACGGGGCC AGGGAGCTCACAACACTATGGAGCCATGAGAACCCTGACTGAACAGTATGAGGAGGTGGACCA GTTTGGGAACACAGTCTCATGTCTTCCACCAGTACCGAGCAGGCAGAGCCACCCAGGAA CCCAGGCTCCACCTCGGGTCCACGCCTCCCCCTTGCTGAGGCAGTTCTGTCAGAGCCAGC TGGGTTGAGCAGTGACCTGACAGAAGCTGAGACGGTGCAGGTGTCTGCAGTACTCCAGCC AGCTGCCAGCTCGAGGGC	
	ORF Start: at 2	ORF Stop: end of sequence
	SEQ ID NO: 80	405 aa MW at 43320.7kD
NOV20e, 249263166 Protein Sequence	TKLSGEQPMEGSHQGAPESPDLSQRNQKELQGLLNQVQALEKEAASSVDVQALRRLFEAVPQL GGAAPQAPAAHQKPEASVEQAFGELTRVSTVEVAQLKEQTLARLLDIEAVHKALSSMSLQPE ASARGHFQGPDKDHAHKISVTVSSSARPSGSGQEVGGQTAVKNQAKVECHTEAQSQVKIRNH TEARGHTASTAPSTRRQETSREYLCPPRVLPSRDSPPSSPTFISIQSATRKPLETPSFKNPD VSVKSTQLAQDIGQALLHQKGVQDKTITQCSVQPEPAPPSASPLPRGWQKSVLELQGTGP GSSQHYGAMRTVTEQYEEVDQFGNTVLMSSTTVTEQAEPNPNPGSHLGLHASPLLRQFLHSPA GFSSDLTEAETVQVSCSYSQPAAQLEG	
	SEQ ID NO: 81	1216 bp

NOV20f, 249263170 DNA Sequence	CACCAAGCTTTCAGGAGAGCAGCCCATGGAAGGTTCCCAACCAAGGGGCCCTGAGAGCCCTGA CAGTCTGCAAAGAAACCAGAAAGAGCTCCAGGGCCTCCTGAACCAGGTGCAAGCCCTGGAGAA GGAGGCCGCAAGCAGTGTGGACGTGCAGGCCCTGCGGAGGCTCTTTGAGGCCGTGCCCCAGCT GGGAGGGGCTGCTCCTCAGGCTCCTGCTGCCACCAAAGCCGAGGCTCAGTGGAGCAGGC CTTTGGGGAGCTGACACGGGTGACACGGAAGTTGCTCAACTGAAGGAACAGACCTTGGCAAG GCTGCTGGACATTGAAGAGGCTGTGCACAAGGCACTCAGTCCATGTCTAGCCTCCAGCCTGA GGCCAGTGCCAGAGGCCATTTCAGGGACCTCCAAAAGACCAGTGCCCAAGATCAGTGT CACAGTCAGCAGTAGCGCCAGGCCAGTGGCTCAGGCCAGGAGGTGCGAGGTCAAACCTGTAGT CAAGAACCAAGCCAAGGTTGAATGCCACACTGAGGCCAGAGTCAAGTCAAGATCAGAAATCA CACAGAGGCCAGAGGTACACAGCCTCAACTGCCCTTCCACCAGGAGGAGGAGACATCAAG AGAGTATTGTGCCCTCCTCGGGTTTACCTTCCAGCCGAGATTCTCCCTCCTCCCAACATT TATCTCCATCCAGTGGGCCACAAGGAGCCTCTAGAGACTCCAGCTTTAAGGGCAACCTGA TGTCTCAGTGAAGACACAACTGGCTCAGGACATAGGCCAGGCCCTGCTCCACCAGAAAGG TGTCCAAGACAAAACCTGGGAAGAAGGACATACCCAGTGTCTGTGCAACCTGAACCTGCCCC TCCCTCAGCCAGTCCCCTGCCAGAGGGTGGCAAAAGAGTGTCTGGAGCTACAGACGGGGCC AGGGAGCTCACAACACTATGGAGCCATGAGAACCGTGAACAGTATGAGGAGGTGGACCA GTTTGGGAACACAGTCTCATGTCTTCCACACAGTCACCGAGCAGGAGGCCACCCAGGAA CCCAGGCTCCACCTCGGGCTCCAGCCTCCCCCTTGCTGAGGCAGTTCTGCACAGCCAGC TGGGTTACAGCAGTGACCTGACAGAAGCTGAGACGGTGCAGGTGCTCCTGCAGCTACTCCAGCC AGCTGCCAGCTCGAGGGC		
	ORF Start: at 2	ORF Stop: end of sequence	
	SEQ ID NO: 82	405 aa	MW at 43348.7kD
NOV20f, 249263170 Protein Sequence	TKLSGEQPMESHQAPESPDLSLRNQKELQGLLNQVQALEKEAASSVDVQALRRLFEAVPQL GGAAPQAPAAHQKPEASVEQAFGLTRVSTEVAQLKEQTLARLLDIEEAVHKALSSMSLQPE ASARGHFQPPKDHSAHKISVTVSSSARPSGSGQEVGGQTVVKNQAKVECHTEAQSQVKIRNH TEARGHTASTAPSTRRQETSREYLCPPRVLPSSRDSPPSPFTFISIQSATRKPLETPSFKGNPD VSVKSTQLAQDIGQALLHQKGVQDKTGKKDITQCSVQPEPAPPSASPLPRGWQKSVLELQGTG GSSQHYGAMRTVTEQYEEVDQFGNTVLMSSSTTVTEQAEPNPNPGLHSLHPLRLQFLHSPA GFSSDLTEAETVQVSCSYSPAQLEG		
	SEQ ID NO: 83	2115 bp	
NOV20g, CG111501-03 DNA Sequence	CAAGAAGGTGCTGTTGGAGCCAGCAGAACAGAACCAATTGAACAGAACCTCCAGAGGAAC GACGAACCCTGAGACCACAGCTGCTACAGACCACAAACACCCCATCAGCCAAGAGAGACCTT GCTGCTGTGACAGGACCTGACTTTCAGCTGGAGGCCACCGTGTGAGGACTCCATCCAGCAA GCCTCTGAGCCCCCTGAAGGACCCCTTCTTCACTCCACAGCAGCCCTGCTGGCCAGAGAACC CTGGAGGGTCACAGACAAAGACCCCAAACTGGACCCCAACATGCCCCCAAGAAGAAGCCG CAGCTGCCCCCTAAACCTGCACACCTAACCCAGAGCCACCCTCCTCAGAGGCTGCCCAAGCCC TTGCCTCTATCTCCAGCTTTTCTCGGAGGTGGGGCAAAGAGAACCAACGAGGTGAGAGA GATACAGCCATCCCTCAGCCAGCCAAGGTTCCCACTACTGTAGACCGAGGCCACATACCTCTG GCCAGATGTCCAGTGGACATAGCCAGCCAGCTTACAACATGCCCTCAGCACCACGGCCCCC AGGCCCAAGAATCAGGTACAGGCAGCAATGCCAGAGCTCTGAGCCCCCAAGTTCATAT GCCCTCAACCATGATCCACCTCACCACAGTGGGGCCCCGGCCCCCTCAGGAGAGCAGCCCATG GAAGGTTCCCAACCAAGGGGCCCTGAGAGCCCTGACAGTCTGCAAAGAAACCAGAAAGAGCTC CAGGGCCTCCTGAACCAGGTGCAAGCCCTGGAGAAGGAGGCCGCAAGCAGTGTGGACGTGCAG GCCCTGCGGAGGCTCTTTGAGGCCGTGCCCAGCTGGGAGGGGCTGCTCCTCAGGCTCCTGCT GCCCACCAAAAGCCGAGGCCTCAGTGGAGCAGGCCTTTGGGAGCTGACACGGGTGACACG GAAGTTGCTCAACTGAAGGAACAGACCTTGGCAAGGCTGTGAGACATTGAAGAGGCTGTGCAC AAGGCACTCAGTCCATGTCTAGCTCCAGCCTGAGGCCAGTGCCAGAGGCCATTTCAGGGA CCTCCAAAAGACCACAGTGCCCAAGATCAGTGTACAGTCAGCAGTAGCGCCAGGCCAGT GGCTCAGGCCAGGAGGTGAGGTTCAAACTGCAGTCAAGAACCAAGCCAAGGTTGAATGCCAC ACTGAGGCCAGAGTCAAGTCAAGATCAGAAATCACACAGAGGCCAGAGGTACACAGCCTCA ACTGCCCTTCCACCAGGAGGCAGGAGACATCAAGAGAGTATTTGTGCCCTCCTCGGGTTTTA CCTTCCAGCCGAGATTCTCCCTCCTCCCAACATTATCTCCATCCAGTGGGCCACAAGGAAG CCTCTAGAGACTCCAGCTTTAAGGGCAACCTGATGTCTCAGTGAAGAAGCACACAACCTGGCT CAGGACATAGGCCAGGCCTGCTCCACCAGAAAGGTGTCCAAGACAAAACCTGGGAAGAAGGAC ATCACCAGTGTCTGTGCAACCTGAACCTGCCCTCCTCAGCCAGTCCCTGCCAGAGGG TGGCAAAGAGTGTCTGGAGCTACAGACGGGGCCAGGGAGCTCAACAACATATGGAGCCATG AGAACCGTGAACAGTATGAGGAGGTGGACAGTTTGGGAACACAGTCTCATGTCTTCC ACCACAGTACCGAGCAGGCAGAGCCACCCAGGAACCCAGGCTCCACCTCGGGCTCCAGGCC TCCCCCTGTGTAGGCAGTTCCTGCACAGCCAGCTGGGTTAGCAGTGAACCTGACAGAAGCT GAGACGTTGACAGTGTCTGACGTAATCCAGCCAGTCCCAAGTGAAGGCCACCGCTCCC ACCACACCTGCCACCTGTTCTGGCCTCCACTGCCCCAGGACTGAAGTGGGTACCTGCCTCCT GTACACTGGAGCAAGGACCAAGAGGAAATGGCATCTTCAGAGGATTACTGTGGGCCATTTCCTC TTTGCGAGTTCTTCAATAGGCCAGTTCTTCCAAATGGAAAAGAAAGGTCTGGAAGAGGCC CACAGAGTTGCACAGGCTGGGGTAGGATGGGGC		

	ORF Start: ATG at 295		ORF Stop: TGA at 1873
	SEQ ID NO: 84	526 aa	MW at 56300.1kD
NOV20g. CG111501-03 Protein Sequence	MPPKKKQQLPPKPAHLTQSHPPQRLPKPLPLSPSFSSEVGQREHQRGERTAI PQPAKVPTTV DRGHIPLARCPSGHSQPSLQHGLSTTAPRPTKNQATGSNAQSSEPPKLNALNHDP TSPQWGPG PSGEQPMEGSHQGAPESPD SLQRNQKELQGLLNQVQALEKEAASSVDVQALRRLFEAVPQLGG AAPQAPAAHQKPEASVEQAFGELTRVSTEVAQLKEQTLARLLDIEEAVHKALSSMSSLQPEAS ARGHFQGP PKDHS AHKISVTVSSSARPSGSGQEVGGQTAVKNQAKVECHTEAQSQVKIRNHT E ARGHTASTAPSTRRQETSREYLCPPRVLPSSRDS PSSPTFISIQSATRKPLETPSFKGNPDVS VKSTQLAQDIGQALLHQKGVQDKTGKKDITQCSVQPEPAPPSASPLPRGWQKSVLELQTGPGS SQHYGAMRTVTEQYEEVDQFGNTVLMSSTTVTEQAEPPRNPGSHLGLHASPLL RQFLHSPAGF SSDLTEAETVQVSCSYSQPAAQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 20B.

Table 20B. Comparison of NOV20a against NOV20b through NOV20g.		
Protein Sequence	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV20b	1..979 1..979	893/979 (91%) 894/979 (91%)
NOV20c	1..600 4..603	514/600 (85%) 515/600 (85%)
NOV20d	1300..1698 4..402	365/399 (91%) 365/399 (91%)
NOV20e	1300..1698 4..402	366/399 (91%) 366/399 (91%)
NOV20f	1300..1698 4..402	365/399 (91%) 365/399 (91%)
NOV20g	1186..1698 14..526	460/513 (89%) 461/513 (89%)

5

Further analysis of the NOV20a protein yielded the following properties shown in Table 20C.

Table 20C. Protein Sequence Properties NOV20a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV20a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 20D.

Table 20D. Geneseq Results for NOV20a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABP06536	Human ORFX protein sequence SEQ ID NO:13054 - Homo sapiens, 165 aa. [WO200192523-A2, 06-DEC-2001]	406..566 2..162	160/161 (99%) 161/161 (99%)	5e-88
AAO05743	Human polypeptide SEQ ID NO 19635 - Homo sapiens, 132 aa. [WO200164835-A2, 07-SEP-2001]	202..333 1..132	124/132 (93%) 127/132 (95%)	1e-64
ABP06405	Human ORFX protein sequence SEQ ID NO:12792 - Homo sapiens, 135 aa. [WO200192523-A2, 06-DEC-2001]	1111..1220 28..135	102/110 (92%) 103/110 (92%)	5e-54
ABP33418	Human ORF2391 protein, SEQ ID NO:4782 - Homo sapiens, 113 aa. [WO200190366-A2, 29-NOV-2001]	1299..1410 2..113	108/112 (96%) 108/112 (96%)	7e-53
AAB41480	Human ORFX ORF1244 polypeptide sequence SEQ ID NO:2488 - Homo sapiens, 113 aa. [WO200058473-A2, 05-OCT-2000]	1299..1410 2..113	108/112 (96%) 108/112 (96%)	7e-53

5

In a BLAST search of public sequence databases, the NOV20a protein was found to have homology to the proteins shown in the BLASTP data in Table 20E.

Table 20E. Public BLASTP Results for NOV20a				
Protein Accession Number	Protein/Organism/Length	NOV20a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O70373	Xin - Mus musculus (Mouse), 1677 aa.	1..1698 1..1677	1069/1734 (61%) 1217/1734 (69%)	0.0
Q8TCG7	Hypothetical 112.1 kDa protein - Homo sapiens (Human), 1059 aa (fragment).	568..1327 4..724	606/772 (78%) 628/772 (80%)	0.0

BAC04783	CDNA FLJ39102 fis, clone NTONG2002948, moderately similar to Mus musculus Xin mRNA - Homo sapiens (Human), 526 aa.	1173..1698 1..526	525/526 (99%) 526/526 (99%)	0.0
BAC04655	CDNA FLJ38622 fis, clone HEART2008364, moderately similar to Mus musculus Xin mRNA - Homo sapiens (Human), 394 aa.	1305..1698 1..394	394/394 (100%) 394/394 (100%)	0.0
Q91957	XIN - Gallus gallus (Chicken), 2562 aa.	20..746 17..794	398/793 (50%) 520/793 (65%)	0.0

Pfam analysis predicts that the NOV20a protein contains the domains shown in the Table 20F.

Table 20F. Domain Analysis of NOV20a			
Pfam Domain	NOV20a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 21.

The NOV21 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 21A.

Table 21A. NOV21 Sequence Analysis			
	SEQ ID NO: 85	618 bp	
NOV21a, CG112595-01 DNA Sequence	GCCCTTCCATGCCAGACCTCAGCAAGTGGTCCGGGCCCTTGAGCCTGCAAGAAGTGGACGAGC AGCCGCAGCACCCGCTGCATGTACCTACGCCGGGCGGCGGTGGACGAGCTGGGCAAAGTGC TGACGCCCACCCAGGTTAAGAATAGACCCACGACATTTCTGGGATGGTCTTGATTCAGGGA AGCTCTACACCTTGGTCCTGACAGACCCGGATGCTCCAGCAGGAAGGATCCCAAATACAGAG AATGGCATCATTTCCTGGTGGTCAACATGAAGGGCAATGACATCAGCAGTGGCACAGTCCTCT CCGATTATGTGGGCTCGGGGCCCTCCAAAGGGCACAGGCCCTCCACCGCTATGTCTGGCTGGTTT ACGAGCAGGACAGGCCGCTAAAGTGTGACGAGCCATCCTCAGCAACCGATCTGGAGACCACC GTGGCAAATTCAGGTGGCGTCTTCCGTAAAAAGTATGAGCTCAGGGCCCCGGTGGCTGGCA CGTGTTACCAGGCCGAGTGGGATGACTATGTGCCAAACTGTACGAGCAGCTGTCTGGGAAGT AGGGGGTTAGCTTGGGGACCTGAAGTGTCTGGAGGCCCCACCACACTCTG		
	ORF Start: ATG at 9		ORF Stop: TAG at 567
	SEQ ID NO: 86	186 aa	MW at 20957.5kD
NOV21a, CG112595-01 Protein Sequence	MPDLKWSGPLSLQEVDEQPQHPLHVTYAGAAVDELGKVLTPQTQVKNRPTSISWDGLDSGLY TLVLTDPDAPSRKDPKYREWHHFLVNMKGNDISSGTVLSYVSGSPPKGTGLHYVWLVEQ DRPLKCDEPILSNRSGDHRGKFKVASFRKKYELRAPVAGTCYQAEWDDYVPKLYEQLSGK		
	SEQ ID NO: 87	1434 bp	
NOV21b, CG112595-02 DNA Sequence	GAGCCAGTGTGCTGAGCTCTCCGCTCGCCTCTGTCGCCCGCGCCTGGCCTACCGGGCACTC CCGGCTGCACGCTCTGCTTGGCTCGCCATGCCGGTGGACCTCAGCAAGTGGTCCGGGCCCTT GAGCCTGCAAGAAGTGGACGAGCAGCCGAGCACCCGCTGCATGTACCTACGCCGGGCGGC		

	GGTGGACGAGCTGGGCAAAGTGTGACGCCCACCCAGGTTAAGAATAGACCCACCAGCATTTC GTGGGATGGTCTTGATTACAGGAAGCTCTACACCTTGGTCCTGACAGACCCGGATGCTCCCAG CAGGAAGGATCCCAAATACAGAGAATGGCATCATTTCTGGTGGTCAACATGAAGGGCAATGA CATCAGCAGTGGCACAGTCTCTCCGATTATGTGGGCTCGGGGCCTCCCAAGGGCACAGGCCT CCACCGCTATGTCTGGCTGGTTACGAGCAGGACAGGCCGCTAAAGTGTGACGAGCCCATCCT CAGCAACCGATCTGGAGACCACCGTGGCAAATCAAGGTGGCGTCTTCCGTAAAAAGTATGA GCTCAGGGCCCCGGTGGCTGGCACGTGTACCAGGCCAGTGGGATGACTATGTGCCCAAACCT GTACGAGCAGCTGTCTGGGAAGTAGGGGGTTAGCTTGGGGACCTGAACCTGTCTGGAGGCCCC AAGCCATGTTCCCCAGTTCAGTGTTCATGTATAATAGATTTCTCCTTCTGCCCCCTTG GCATGGGTGACACCTGACCAGTCAGATGGTAGTTGAGGGTGACTTTCTGCTGCCTGGCCTT TATAATTTTACTCACTCACTCTGATTATGTTTGTATCAAATTTGAACCTTCATTTTGGGGGT ATTTTGGTACTGTGATGGGGTCATCAAATTATTAATCTGAAAATAGCAACCCAGAATGTAAAA AAGAAAAGCTGGGGGAAAAAGACCAGGTCTACAGTGATAGAGCAAAGCATCAAAGAATCTT TAAGGGAGGTTTAAAAAAGAAAAAAGATTGGTTGCCTCTGCCTTTGTGATCGCTGA GTCCAGAATGGTACACAATGTGATTTTATGGTGATGTCACTCACCTAGACAACCCAGAGGCTGG CATTGAGGCTAACCTCCAACACAGTGCATCTCAGATGCCTCAGTAGGCATCAGTATGTCACTC TGGTCCCTTTAAAGAGCAATCCTGGAAGAAGCAGGAGGGAGGGTGGCTTTGCTGTTGTTGGGA CATGGCAATCTAGACGGCAGCAGCGCTCGTGACAGCTTGGGAGGAAACCTGAGATCTGTGT TTTTAAATTGATCGTCTTTCATGGGGTAAGAAAAGCTGGTCTGGAGTTGCTGAATGTTGCA TTAATTGTGCTGTTTCTTGTAGTTGAATAAAAATAGAAACCTGAATG		
	ORF Start: ATG at 92		ORF Stop: TAG at 653
	SEQ ID NO: 88	187 aa	MW at 21056.6kD
NOV21b, CG112595-02 Protein Sequence	MPVDLSKWSGPLSLQEVDQPHPLHVTYAGAAVDELGKVLTPQVKNRPTSI SWDGLDSGKL YTLVLTDPDAPSRKDPKYREWHHFLVVMKGNDISSGTVLSDYVSGPPKGTGLHRYVWLVE QDRPLKCEPILSNRSGDHRGKFKVASFRKKYELRAPVAGTCYQAEWDYVPKLYEQLSGK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 21B.

Table 21B. Comparison of NOV21a against NOV21b.		
Protein Sequence	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV21b	3..186	184/184 (100%)
	4..187	184/184 (100%)

5

Further analysis of the NOV21a protein yielded the following properties shown in Table 21C.

Table 21C. Protein Sequence Properties NOV21a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3603 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV21a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 21D.

Table 21D. Geneseq Results for NOV21a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE21677	Human phosphoethanolamine binding protein (PEBP) - Homo sapiens, 187 aa. [WO200218623-A2, 07-MAR-2002]	3..186 4..187	184/184 (100%) 184/184 (100%)	e-109
AAR49943	Human hippocampal cholinergic neurotrophic peptide precursor - Homo sapiens, 187 aa. [WO9405788-A, 17-MAR-1994]	3..186 4..187	184/184 (100%) 184/184 (100%)	e-109
AAR27718	HCNP precursor protein #2 - Homo sapiens, 187 aa. [EP511816-A, 04-NOV-1992]	3..186 4..187	184/184 (100%) 184/184 (100%)	e-109
AAR64268	Phosphatidylethanolamine binding protein - Homo sapiens, 187 aa. [EP628631-A, 14-DEC-1994]	3..186 4..187	183/184 (99%) 183/184 (99%)	e-109
AAE21676	Mouse phosphoethanolamine binding protein (PEBP) - Mus musculus, 187 aa. [WO200218623-A2, 07-MAR-2002]	3..186 4..187	160/184 (86%) 170/184 (91%)	1e-95

5

In a BLAST search of public sequence databases, the NOV21a protein was found to have homology to the proteins shown in the BLASTP data in Table 21E.

Table 21E. Public BLASTP Results for NOV21a				
Protein Accession Number	Protein/Organism/Length	NOV21a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH31102	Prostatic binding protein - Homo sapiens (Human), 187 aa.	3..186 4..187	184/184 (100%) 184/184 (100%)	e-109

P30086	Phosphatidylethanolamine-binding protein (PEBP) (Neuropolypeptide h3) (Hippocampal cholinergic neurostimulating peptide) (HCNP) (Raf kinase inhibitor protein) (RKIP) - Homo sapiens (Human), 186 aa.	3..186 3..186	184/184 (100%) 184/184 (100%)	e-109
S46485	phosphatidylethanolamine-binding protein - crab-eating macaque, 187 aa.	3..186 4..187	180/184 (97%) 181/184 (97%)	e-106
P48737	Phosphatidylethanolamine-binding protein (PEBP) - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 186 aa.	3..186 3..186	180/184 (97%) 181/184 (97%)	e-106
P13696	Phosphatidylethanolamine-binding protein (PEBP) (Basic cytosolic 21 kDa protein) - Bos taurus (Bovine), 186 aa.	3..186 3..186	173/184 (94%) 177/184 (96%)	e-102

PFam analysis predicts that the NOV21a protein contains the domains shown in the Table 21F.

Table 21F. Domain Analysis of NOV21a			
Pfam Domain	NOV21a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PBP	1..171	91/201 (45%) 162/201 (81%)	1.1e-84

5

Example 22.

The NOV22 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 22A.

Table 22A. NOV22 Sequence Analysis			
	SEQ ID NO: 89	662 bp	
NOV22a, CG112624-01 DNA Sequence	CGCAGCCAGCACCGGGCGGAGAGGGCTACCATGGGGAAAATCGCGCTGCAACTCAAAGCCACG CTGGAGAACATCACCAACCTCCGGCCCCGTGGGCGAGGACTTCCGGTGGTACCTGAAGGACAGT GTGGCACTGAAGGGGGCCGTGGCAGTGCTTCCATGGTCCAGAAGTGCAAGCTGTGTGCAAGA GAAAATTCCATCGAGATTTTAAGCAGCACCATCAAGCCTTACAATGCTGAAGACAATGAGAAC TTCAAGACAATAGTGGAGTTTGAGTGCCGGGGCCTTGAACCAAGTTGATTTCAGCCCGCAGGCT GGGTTTGCTGCTGAAGGTGTGGAGTCAGGGACAGCCTTCAGTGACATTAATCTGCAGGAGAAG GACTGGACTGACTATGATGAAAAGGCCAGGAGTCTGTGGGAATCTATGAGGTCACCCACCAG TTTGTGAAGTGCTGATCCCTCTTCTTCCAGTTGCCCTTAAGAACTGAGAAAGGACAAAGTA CTCTAAGCAGCAGAGCCACAGAGGCTCGTTTCTTTGACCCTTGTCTCCTGGTGGCTATACGA AACCTTCACAATCTGCATGCTGGACTTTATTACAGCTTCCAAGCCCCATCAATAAGGCCCT GTTACGCTGCACTGGTGCATGAAGGTGAAAT		
	ORF Start: ATG at 31		ORF Stop: TGA at 454
	SEQ ID NO: 90	141 aa	MW at 15790.6kD

NOV22a, CG112624-01 Protein Sequence	MGKIALQLKATLENITNLRPVGEDFRWYLKDSVALKGGRGSASMVQKCKLCARENSIETLSST IKPYNAEDNENFKTIVEFECRGLPEVDFQPQAGFAAEGVESGTAFSDINLQEKDWDYDEKAQ ESVGIYEVTHQFVKC
--	--

Further analysis of the NOV22a protein yielded the following properties shown in Table 22B.

Table 22B. Protein Sequence Properties NOV22a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV22a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 22C.

Table 22C. Geneseq Results for NOV22a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG89328	Human secreted protein, SEQ ID NO: 448 - Homo sapiens, 160 aa. [WO200142451-A2, 14-JUN-2001]	1..141 1..160	117/160 (73%) 121/160 (75%)	2e-58
AAG01632	Human secreted protein, SEQ ID NO: 5713 - Homo sapiens, 107 aa. [EP1033401-A2, 06-SEP-2000]	1..85 1..104	85/104 (81%) 85/104 (81%)	4e-40
ABB60841	Drosophila melanogaster polypeptide SEQ ID NO 9315 - Drosophila melanogaster, 161 aa. [WO200171042-A2, 27-SEP-2001]	1..140 1..160	63/160 (39%) 91/160 (56%)	2e-28
AAG48660	Arabidopsis thaliana protein fragment SEQ ID NO: 61473 - Arabidopsis thaliana, 167 aa. [EP1033405-A2, 06-SEP-2000]	1..126 1..151	50/152 (32%) 78/152 (50%)	2e-15
AAG17798	Arabidopsis thaliana protein fragment SEQ ID NO: 18956 - Arabidopsis thaliana, 167 aa. [EP1033405-A2, 06-SEP-2000]	1..126 1..151	50/152 (32%) 78/152 (50%)	2e-15

10

In a BLAST search of public sequence databases, the NOV22a protein was found to have homology to the proteins shown in the BLASTP data in Table 22D.

Table 22D. Public BLASTP Results for NOV22a				
Protein Accession Number	Protein/Organism/Length	NOV22a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NWX4	CDNA FLJ20580 fis, clone REC00516 (Similar to hypothetical protein FLJ20580) - Homo sapiens (Human), 160 aa.	1..141 1..160	141/160 (88%) 141/160 (88%)	1e-74
Q9DCH5	0610037L13Rik protein (RIKEN cDNA 0610037L13 gene) - Mus musculus (Mouse), 196 aa.	3..141 39..196	132/158 (83%) 136/158 (85%)	6e-70
Q9D1H2	I110008H16Rik protein - Mus musculus (Mouse), 168 aa.	1..133 1..154	114/154 (74%) 122/154 (79%)	9e-56
T22286	hypothetical protein F46B6.3 - Caenorhabditis elegans, 491 aa.	4..138 328..482	65/155 (41%) 95/155 (60%)	1e-29
AAF58406	CG4646-PA - Drosophila melanogaster (Fruit fly), 163 aa.	1..140 1..160	63/160 (39%) 91/160 (56%)	6e-28

- 5 PFam analysis predicts that the NOV22a protein contains the domains shown in the Table 22E.

Table 22E. Domain Analysis of NOV22a			
Pfam Domain	NOV22a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 23.

- 10 The NOV23 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 23A.

Table 23A. NOV23 Sequence Analysis			
	SEQ ID NO: 91	1019 bp	
NOV23a, CG113823-01 DNA Sequence	GCAGTGAAGCTCTGGGAAATCCTTCATTAAATCATTACACCTTCAGGGACATTTAAGAACTCACA ATGGAGAAAGTCTCCATGAATGGAAGGAATGTGGGAGAGGCTTTATTCATCCACAGACCTTG CTGTGCGTATACAACTCACAGGTCAGAAAAACCTACAAATGTAAGGAATGTGGAAAAGGAT TTAGATATTCTGCATACCTTAATATTACATGGGAACCCACACTGGAGACAATCCCTATGAGT		

	GTAAGGAGTGTGGGAAAGCCTTCACCAGGTCTGTCAACTTACTCAGCACAGAAAACTCACA CTGGAGAGAAACCTTATAAATGTAAGGATTGTGGGAGAGCCTTCACTGTTTCCTCTTGCTTAA GTCAACATATGAAAAATCCATGTGGGTGAGAAGCCTTATGAATGCAAGGAATGTGGGATAGCCT TCACTAGATCTTCTCAACTTACTGAACATTTAAAACTCAGCTGCAAAGGATCCCTTTGAAT GTAAGATATGTGGAAAATCCTTTAGAAATTCCTCATGCCTCAGTGATCACTTTTCAATTCACA CTGGAATAAAACCTTATAAATGTAAGGATTGTGGGAAAGCCTTCACTCAGAACTCAGACCTTA CTAAGCATGCACGAACTCAGAGTGGAGAGAGGCCCTATGAATGTAAGGAATGTGGAAAGGCCT TTGCCAGATCCTCTCGCCTTAGTGAACATACAAGAACTCAGCTGGAGAGAAGCCTTTTGAAT GTGTCAAATGTGGGAAAGCCTTTGCTATTTCTTCAAATCTTAGTGGACATTTGAGAATTCACA CTGGAGAGAAGCCTTTGAGTGCCTGGAATGTGGTAAAGCATTACGCATTCTCCAGTCTTA ATAATCACATGCGGACCCACAGCGCCAAAAAACCATTCAGTGTATGGAATGTGGCAAAGCCT TTAAGTTTCCCACGTGTGTTAACCTTCACATGCGGATTCACACTGGAGAAAACCTTACAAATGT AACAGTGTGGA		
	ORF Start: ATG at 219		ORF Stop: TAA at 1008
	SEQ ID NO: 92	263 aa	MW at 29620.6kD
NOV23a, CG113823-01 Protein Sequence	MGHTHTGDNPYECKECGKAFTSRQLTQHRKHTHTGEKPYKCKDCGRAFTVSSCLSQHMKIHVGE KPYECKECGIAFTRSSQLTEHLKHTAKDPFECKICGKSFRNSSCLSDFRIHTGKPYKCKD CGKAFTQNSDLTKHARTHSGERPYECKECGKAFTSRRLSEHTRHTHTGEKPFECVKCGKAFAI SSNLSGHLRIHTGEKPFECLECGKAFTHSSSLNNHMRTHSAKKPFTCMCEGKAFKFPPTCVNLH MRIHTGENPTM		
	SEQ ID NO: 93	894 bp	
NOV23b, CG113823-02 DNA Sequence	GCCCTTTGTGCGTATACAAACTCACAGGTCAGAAAAACCTTACAAATGTAAGGAATGTGGAAA AGGATTTAGATATTCTGCATACCTTAATATTCACATGGGAACCCACACTGGAGACAATCCCTA TGAGTGTAGGAGTGTGGGAAAGCCTTCACCAGGTCTGTCAACTTACTCAGCACAGAAAAAC TCACACTGGAGAGAAACCTTATAAATGTAAGGATTGTGGGAGAGCCTTCACTGTTTCCTCTTG CTTAAGTCAACATATGAAAATCCATGTGGGTGAGAAGCCTTATGAATGCAAGGAATGTGGGAT AGCCTTCACTAGATCTTCTCAACTTACTGAACATTTAAAACTCAGCTGCAAAGGATCCCTT TGAATGTAAGATATGTGGAAAATCCTTTAGAAATTCCTCATGCCTCAGTGATCACTTTTGAAT TCACACTGGAATAAAACCTTATAAATGTAAGGATTGTGGGAAAGCCTTCACTCAGAACTCAGA CCTTACTAAGCATGCACGAACTCAGAGTGGAGAGAGGCCCTATGAATGTAAGGAATGTGGAAA GGCCTTTGCCAGATCCTCTCGCCTTAGTGAACATACAAGAACTCAGCTGGAGAGAAGCCTTT TGAATGTGTCAAATGTGGGAAAGCCTTTGCTATTTCTTCAAATCTTAGTGGACATTTGAGAAT TCACACTGGAGAGAAGCCTTTGAGTGCCTGGAATGTGGTAAAGCATTACGCATTCTCCAG TCTTAATAATCACATGCGGACCCACAGCGCCAAAAAACCATTCAGTGTATGGAATGTGGCAA AGCCTTTAAGTTTCCCACGTGTGTTAACCTTCACATGCGGATCCACACTGGAGAAAACCTTAC AATGTAACAGTG		
	ORF Start: ATG at 98		ORF Stop: TAA at 887
	SEQ ID NO: 94	263 aa	MW at 29620.6kD
NOV23b, CG113823-02 Protein Sequence	MGHTHTGDNPYECKECGKAFTSRQLTQHRKHTHTGEKPYKCKDCGRAFTVSSCLSQHMKIHVGE KPYECKECGIAFTRSSQLTEHLKHTAKDPFECKICGKSFRNSSCLSDFRIHTGKPYKCKD CGKAFTQNSDLTKHARTHSGERPYECKECGKAFTSRRLSEHTRHTHTGEKPFECVKCGKAFAI SSNLSGHLRIHTGEKPFECLECGKAFTHSSSLNNHMRTHSAKKPFTCMCEGKAFKFPPTCVNLH MRIHTGENPTM		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 23B.

Table 23B. Comparison of NOV23a against NOV23b.		
Protein Sequence	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV23b	1..263	263/263 (100%)
	1..263	263/263 (100%)

Further analysis of the NOV23a protein yielded the following properties shown in Table 23C.

Table 23C. Protein Sequence Properties NOV23a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3008 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV23a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 23D.

Table 23D. Geneseq Results for NOV23a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB43912	Human cancer associated protein sequence SEQ ID NO:1357 - Homo sapiens, 580 aa. [WO200055350-A1, 21-SEP-2000]	1..261 262..522	260/261 (99%) 260/261 (99%)	e-164
AA Y73346	HTRM clone 619699 protein sequence - Homo sapiens, 549 aa. [WO9957144-A2, 11-NOV-1999]	1..261 231..491	259/261 (99%) 260/261 (99%)	e-164
AAB94123	Human protein sequence SEQ ID NO:14372 - Homo sapiens, 243 aa. [EP1074617-A2, 07-FEB-2001]	1..232 1..232	232/232 (100%) 232/232 (100%)	e-144
ABG05639	Novel human diagnostic protein #5630 - Homo sapiens, 1560 aa. [WO200175067-A2, 11-OCT-2001]	4..261 1173..1430	159/258 (61%) 192/258 (73%)	e-100
ABG01726	Novel human diagnostic protein #1717 - Homo sapiens, 1342 aa. [WO200175067-A2, 11-OCT-2001]	4..261 1060..1317	152/258 (58%) 199/258 (76%)	1e-98

- 10 In a BLAST search of public sequence databases, the NOV23a protein was found to have homology to the proteins shown in the BLASTP data in Table 23E.

Table 23E. Public BLASTP Results for NOV23a				
Protein Accession Number	Protein/Organism/Length	NOV23a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96SX9	CDNA FLJ14574 fis, clone NT2RM4000751, moderately similar to zinc finger protein 184 - Homo sapiens (Human), 243 aa.	1..232 1..232	232/232 (100%) 232/232 (100%)	e-144
Q8WVX1	Hypothetical 29.9 kDa protein - Homo sapiens (Human), 263 aa.	57..261 1..205	204/205 (99%) 204/205 (99%)	e-125
Q8TAQ5	Similar to unnamed protein product - Homo sapiens (Human), 688 aa.	4..261 301..558	159/258 (61%) 192/258 (73%)	e-100
BAC04086	CDNA FLJ35863 fis, clone TESTI2007524, moderately similar to HYPOTHETICAL ZINC FINGER PROTEIN - Homo sapiens (Human), 475 aa.	4..261 192..449	158/258 (61%) 194/258 (74%)	e-100
Q96MM9	CDNA FLJ32133 fis, clone PEBLM2000308, moderately similar to zinc finger protein 135 - Homo sapiens (Human), 338 aa.	4..261 35..292	155/258 (60%) 192/258 (74%)	1e-98

PFam analysis predicts that the NOV23a protein contains the domains shown in the Table 23F.

Table 23F. Domain Analysis of NOV23a			
Pfam Domain	NOV23a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C2H2	10..32	11/24 (46%) 22/24 (92%)	4.4e-06
zf-C2H2	38..60	11/24 (46%) 20/24 (83%)	7.1e-06
zf-C2H2	66..88	12/24 (50%) 21/24 (88%)	7.5e-07
zf-C2H2	94..116	11/24 (46%) 19/24 (79%)	8.1e-05
zf-BED	79..117	14/52 (27%) 23/52 (44%)	0.58
zf-C2H2	122..144	14/24 (58%) 22/24 (92%)	4.5e-08

zf-C2H2	150..172	13/24 (54%) 18/24 (75%)	3.3e-06
zf-C2H2	178..200	12/24 (50%) 18/24 (75%)	0.00013
zf-BED	163..201	15/52 (29%) 26/52 (50%)	0.22
zf-C2H2	206..228	11/24 (46%) 22/24 (92%)	2e-07
zf-C2H2	234..256	8/24 (33%) 17/24 (71%)	0.0034

Example 24.

The NOV24 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 24A.

5

Table 24A. NOV24 Sequence Analysis			
	SEQ ID NO: 95	1945 bp	
NOV24a, CG114098-01 DNA Sequence	GACAAGCTTCCTAGCGCTATGACTGTCGTCTCCGTCCCGCAGCGGGAGCCGCTCGTCTCGGGT GGCCGCTTGCGCGCTTGGCTTTCTCCCGAGGTTACTTTGGGGCCCTCCCGATGGTGACC ACGGCTCCGCTCCTTTACCCCGGATCCCGGACCCCGGGCACTGCCCCGACCCTCTTCTCTC CCTCATTTCCTAGGGGAGATGGCCGTGTCTGACCCCCAGCCTCGCGCTCCAGCAGCTCTG CCCAACCGCAGCCTCGCGTGGCGGAGGCACTCCTCGGGCAGCGCGAAGAAGCGCGAAAG AAGAAGGTGCGGGCCAGCCCCGAGGGCAGCTGCCAGCCGCTTCCACCAGTACCAGCAGCAC CGGAGCCCGCGACCGGCCGAGCGGAGCGCAGGAGGTCCCGGGCCCGGCCCGCCCTTGGCC CCGAGTCTGCGAGCCGCGCAGCCGCGAGGGAGGCCAGCCCCGACCTTGCCCCGTGCGGCC GCGGCTCCCGGCCAAACCCCTCAGGAAAGAGGTTTAAAAATCAAAGATGGGAAATCGGAG AAAATTGCCCTTCCCATGGCCAGCTTGTTCATGGTATACACTTGTATGAGCAACCAAAGATA AACAGACAGAAAGCAAATATACTTGGCACTAACCAAGATCACCTCTGCAAAAAGAAATGAA AACAACTTTTGGCAGGATTCTGTTTCATCTGACAGAATTCAGAAGCAGGAAAAAAGCCTTTT AAAAATACCGAGAACATTAAAAATTGCAATTGAAGAAATCAGCATTTCTAACTGAAGTGAGC CAAAAGGAAAATTATGCTGGGGCAAAGTTTAGTGATCCACCTTCTCTAGTGTTCTTCCAAAG CCTCTAGTCACTGGATGGGAAGCACTGTTGAAAATCCAACCAAACAGGGAGCTGATGGCA GTACACTTAAACCCCTCCTCAAAGTTCAAACCTTAGATTTCAGATTTCAGTATGTGTGTAATA CATAATTTTCCCATATCCCTGGACTCTGAGAAAATTGGTACAGAAATGGAATTTGCCTTG TTGCAACATACAATTGCAAAAGATGAGTTTAAAAAATTACATACAAACAGCTTGTATTATATT TTATATTTGTAAATACTGTATACCATGTATTATGTGTATATTGTTTACACTTGAGAGGTATA TTATAGTTTGTATGAAAGTATGTATTTGCCCTGCCACATTGCGAGGTGTTTTGTATATAT ACAATGGATAAATTTAAGTGTGTGCTAAGGCACATGGAAGACCGATTTTATTTGCACAAGGT ACTGAGATTTTTCAGAAACAGCTGTCAAATCTCAAGGTGAAGATCTAAATGTGAACAGTT TACTAATGCACTACTGAAGTTTAAATCTGTGGCACAATCAATGTAAGCATGGGGTTGTTTCT CTAAATGATTGTAACTGAAATTACTGAACAACCTCCTATTCCATTTTGTCTAACTCAAT TTCTGGTTTGTATATATCCATTCCAGCTTAATGCCTCTAATTTAATGCCAACAAAATTGG TTGTAATCAAATTTTAAATAATAAATTTGGCCCCCTTTTAAATAGTCTTGACTCTTT GTGTGTGACTGTTTCTCATGTTTGAATGTGTGACTAGGAGATGATTTTGTGTGGTTGGATTTT TTGACTTCTACTTTACTGGCTGAGTGTGAGCCGCATGCCTGGCCATAATCTACATTTTCTT ACCAGGAGCAGCATTGAGGTTTGTAGCATAGTACTTGACTACTCTAGAGGCTGAGACGGGAG CATCTCTGAGCCTGAGAAGTGAGATTGCAATTGAGCTAGGATCAGGCCACTGCACCTCCAGC CTGGGTAACAGACGCTGTCTCAAAAAAAGGCCAAGAGAAAGTAAGGGAGACAGA		
	ORF Start: ATG at 19		ORF Stop: TAG at 979
	SEQ ID NO: 96	320 aa	MW at 34527.4kD
NOV24a, CG114098-01 Protein	MTVVSVPQREPLVLGGRLAPLGFSSRGYFGALPMVTTAPPPLPRI PDPRALPPTLFLPHFLGG DGPCLTPQPRAPAALPNRSLAVAGGTPRAAPKRRKKKVRASPAGQLPSRFHQYQHRSPATG		

Sequence	PSGAQEVPGPAAALAPSPAAAAGTEGASPD LAPLRPAAPGQTPLRKEVLKSKMGKSEKIALPH GQLVHGIHLYEQPKINRQSKYNLPLTKITSAKRNENNFWQDSVSSDRIQKQEKKPFKNTENI KNSHLKKS AFLTEVSQKENYAGAKFSDPPSPSVLPKPPSHWMGSTVENSQNRELM AVHLKTL LKVQT
----------	---

Further analysis of the NOV24a protein yielded the following properties shown in Table 24B.

Table 24B. Protein Sequence Properties NOV24a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.2265 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV24a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 24C.

Table 24C. Geneseq Results for NOV24a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y48508	Human breast tumour-associated protein 53 - Homo sapiens, 211 aa. [DE19813835-A1, 23-SEP-1999]	121..320 15..211	183/200 (91%) 185/200 (92%)	e-101
AA Y48433	Human prostate cancer-associated protein 130 - Homo sapiens, 153 aa. [DE19811194-A1, 16-SEP-1999]	168..320 1..153	153/153 (100%) 153/153 (100%)	3e-85
AAM47280	Human DNA topoisomerase I-15 - Homo sapiens, 139 aa. [WO200181395-A1, 01-NOV-2001]	201..319 21..138	50/121 (41%) 71/121 (58%)	2e-16
AAB42905	Human ORFX ORF2669 polypeptide sequence SEQ ID NO:5338 - Homo sapiens, 139 aa. [WO200058473-A2, 05-OCT-2000]	201..319 21..138	50/121 (41%) 71/121 (58%)	2e-16
AA W40200	Infected cell protein number 4 - Herpes simplex virus, 1297 aa. [WO9804709-A2, 05-FEB-1998]	40..165 670..785	47/126 (37%) 60/126 (47%)	8e-06

10

In a BLAST search of public sequence databases, the NOV24a protein was found to have homology to the proteins shown in the BLASTP data in Table 24D.

Table 24D. Public BLASTP Results for NOV24a				
Protein Accession Number	Protein/Organism/Length	NOV24a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q12796	Proline-rich protein 2 (B4-2 protein) - Homo sapiens (Human), 327 aa.	1..320 1..327	320/327 (97%) 320/327 (97%)	0.0
Q63647	Proline rich protein - Rattus norvegicus (Rat), 269 aa.	34..320 1..269	212/287 (73%) 224/287 (77%)	e-112
Q9NPJ4	CDNA FLJ20312 FIS, clone HEP07362 (HSPC208) (Hypothetical 15.6 kDa protein) (Proline-rich nuclear receptor coactivator 2) - Homo sapiens (Human), 139 aa.	201..319 21..138	50/121 (41%) 71/121 (58%)	7e-16
Q9CR73	0610011E17Rik protein (RIKEN cDNA 0610011E17 gene) - Mus musculus (Mouse), 140 aa.	179..319 1..139	57/144 (39%) 79/144 (54%)	9e-16
Q9CXC6	0610011E17Rik protein - Mus musculus (Mouse), 140 aa.	179..319 1..139	56/144 (38%) 79/144 (53%)	3e-15

- 5 PFam analysis predicts that the NOV24a protein contains the domains shown in the Table 24E.

Table 24E. Domain Analysis of NOV24a			
Pfam Domain	NOV24a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 25.

- 10 The NOV25 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 25A.

Table 25A. NOV25 Sequence Analysis			
	SEQ ID NO: 97	2440 bp	
NOV25a, CG114308-01 DNA	AGTTCTCTGTAGTGTTTGCCAATGTTGGAGCCGTCTGCAAAGTGTCCCCGGCAAGAAGAGGCT GCCTACCACAAGGACTTTAGCTTACTTTTAAAGATTGAAGAAAAAAGAACACAGAAAAAG		

Sequence	AAGAACTCAAAGATACACAAAGTAATTTGAACCAAGGCTCAGAAGTTTTGGAGCCGTGAGGG ATACAGCAGTTTGGTCAATATTGTCTTAACTGCTTCAAATAAATCAGCTTCTCTCAAAGATA AAATGGCAAACCCAAAAGAGAAAAGTGAATGTGTCTGGTAAATGAGTTAGCCCGTTTCAATA GAGTCCAACCCAGTATAAACTTCTGAATGAAAGAGGGCCTGCTCATTCAAAGATGTTCTCAG TGCAGCTGAGTCTTGGTGAGCAGACATGGGAATCCGAAGGCAGCAGTATAAAGAAGGCTCAGC AGGCTGTTGCCAATAAAGCTTTGACTGAATCTACGCTTCCCAAACAGTTTCAAGAAGCCACCA AAAGTAATGTTAAACAATAACCCAGGCAGTATAACTCCAAGTGTGGAAGTGAATGGGCTTGCTA TGAAAAGGGGAGAGCCCTGCCATCTACAGGCCATTAGATCCAAAGCCATTCCCAAATTATAGAG CTAATTACAACCTTTCGGGGCATGTACAATCAGAGGTATCATTGCCAGTGCCCTAAGATCTTTT ATGTTTCACTGCTACTGTAGGAAATAATGAATTTTTTGGGGAAGGAAAGACTCGACAAGCTGCTA GACACAATGTGCAATGAAAGCCCTCAAGCACTGCAGAATGAACCTATTCCAGAAAGATCTC CTCAGAATGGTGAATCAGGAAAGGATATGGATGATGACAAGATGCAATAAGTCTGAGATCA GCTTAGTGTTTGAAATTGCTCTGAAGCGAAATATGCCTGTGAGTTTGGAGTTATTAAAGAAA GTGGACCAACACATATGAAAAGCTTTGTTACTCGAGTGTGAGTAGGAGAGTTCTCTGCAGAAG GAGAAGGAAATAGCAAAAAGTCTTCCAAGAAGCGCGCTGCGACCAACCGCTTACAGGAGCTTA AAAAACTTCCACCTCTTCTGTGGTGGAAAAGCCAAAAGTATTTTTTAAAAAACGCCCTAAAA CAATAGTAAAGGCCGAGCAGAAATATGGCCAAGGGATGAACCTATTAGCCGCTGGCGCAAAA TTCAACAGGCCAAAAGGAAAAGGAGCCGGATTATGTTTTGCTTTCAGAAAAGGAAATGCCTC GACGTGAGAAATTTGTGATGCAAGTGAAGGTAGGCAATGAAGTTGCTACAGGAACAGGACCTA ATAAAAAAGATAGCCAAAAAATGTCTGCAAGCAATGCTGTTACAACCTTGGTTATAAAGCAT CCACTAATCTTCAGGATCAACTTGAGAAGACAGGGGAAAAACAAGGATGGAGTGGTCCAAAGC CTGGGTTTCCTGAACCAACAATAATACTCCAAAAGGAATCTTCATTGTCTCTGTATGTTT ATCAAGAGATGGAAGCCAGCCGCCACAAAGTAATCTCTGGCACTACTCTAGGCTATTGTGAC CCAAAGATATGAACCAACCTTCAAGCTCTTCTTTCAGTATATCTCCCACTCGAATAGTTCAG CTACAATTGCCAGGGAACCTCTTATGAATGGAACATCTTCTACAGCTGAAGCCATAGGTTTAA AAGGAAGTTCTCTACTTCCCTTGTCTCCAGTACAACCTTCAAAACAAGTGAATATTTAG CAAGGATTCAGGCTTTCAGGTTCACTACTGTGATAGACAAAGTGGCAAAGAGTGTGTGACCT GTCTGACATTGACCCCTGTGCAGATGACTTTCATGCTATTGGAAGCTCCATTGAAGCCAGCC ATGATCAGGCAGCCTTAAGTGCCTTGAAACAATTTTCTGAACAAGGACTGGATCCAATCGATG GAGCAATGAATATCGAAAAGGTTCTCTTGAAAAACAAGCCAAGCATCTGAGAGAGAAAGCGG ACAATAACCAAGGCACCCCGGGCTCCATCGCTCAGGACTGCAAGAAATCAAAGTCCGGCCGTCT AGCAGCTCCAGAACCCGGCTGCCACCGCATCTTTATAAACCTGTCAAGCAGCATGAGGGT GTCTGTGTTTCAGGAAATGAATGACTAATACCATTATTTGAGTCTTATGTGAAGACAACACTA TTCTAACACGAGAGATAATATACATGGTACTGTTTATTCCACTGGGGAAAAATAAACTTTGAG CATTCCTGGACTCGAGATCGATCTACTCATGCTGAGCGGAAATGTCCGGTCCGACTAA ATAAGTAGAAATTGAAAAGCAATTACTATTAATAAAAAAACAATTACTTATAACCATCT TTTCTCATTTTTCTTATTACTTATATATTTTAAATATTATTTTTT		
	ORF Start: ATG at 255		ORF Stop: TAG at 2079
	SEQ ID NO: 98	608 aa	MW at 66785.6kD
NOV25a, CGI14308-01 Protein Sequence	MANPKEKTAMCLVNELARFNRPQYKLLNERGPAHSMFVSQSLGLEQTWESEGGSSIKKAQQ AVANKALTESTLKPVPQKPKSNVNNPGSITPTVELNGLAMKRGEPAIYRPLDPKPFPPNYRA NYNFRGMYNQRYHCPVPKIFYVQLTVGNNEFFGEGKTRQARHNAAMKALQALQNEPI PERSP QNGESGKMDDDKDANKSEISLVFEIALKRNPVSFEVIESGPPHMKSFVTRVSVGEFSAEG EGNSKKLSKKRAATTVLQELKLLPLPVVEKPKLFFKKRPKTIKAGPEYQGMNPIISR LAQI QQAKKEKEPDYVLLSERGMPPRRREFVMQVKVNEVATGTGPNKKIAKNAAEAMLLQLGYKAS TNLQDQLEKTGENKWSGPKPGFPEPTNNTPKGILHLSPDVYQEMASRHKVISGTTGLYLSLSP KDMNQPSSSFFSISPTSNSATIARELLMNGTSSTAEI GLKGSSPTPPSPVPQPSKQLEYLA RIQGFQVHYCDRQSGKECVTCLTLAPVQMTFHAIGSSI EASHDQAALSALKQFSEQGLDPIDG AMNIEKGSLEKQAKHLREKADNNQAPPGSIAQDCKKNSAV		
	SEQ ID NO: 99	1899 bp	
NOV25b, CGI14308-02 DNA Sequence	TGTGAGGGATACAGCAGTTTGGTCAATATTGTCTTAACTGCTTCAAATAAATCAGATTGAAG AACTCCCTTCAAGATTTCTGTAGAACAAGTCTGGTGTGATGAAATCCTTCGGCTTTTGTGTT GTCTGGGAAAGACTTTATCTCTCTTCATAATTAAGGATATTTTACCAGCATATACTATTCT AGATGTTCTCAGTGCAGCTGAGTCTTGGTGAGCAGACATGGGAATCCGAAGGCAGCAGTATAA AGAAGGCTCAGCAGGCTGTTGCCAATAAAGCTTTGACTGAATCTACGCTTCCCAAACAGTTT AGAAGCCACCCAAAAGTAATGTTAACAATAACCCAGGCAGTATAACTCCAAGTGTGGAAGTGA ATGGGCTTGCTATGAAAAGGGGAGAGCCTGCCATCTACAGGCCATTAGATCCAAAGCCATTCC CAAATTATAGAGCTAATTACAACCTTTCGGGGCATGTACAATCAGAGGTATCATTGCCAGTGC CTAAGATCTTTTATGTTCACTGCTAGGAAATAATGAATTTTTTGGGGAAGGAAAGACTC GACAAGCTGCTAGACACAATGCTGCAATGAAAGCCCTCCAAGCACTGCAGAATGAACCTATTCT CAGAAAGATCTCCTCAGAATGGTGAATCAGGAAAGGATATGGATGATGACAAAGATGCAATA AGTCTGAGATCAGCTTAGTGTGTTGAAATTGCTCTGAAGCGAAATATGCCCGTCAGTTTGTAGG TTATTAAGAAAGATGGACCACCAACACATATGAAAAGCTTTGTTACTGAGTGTGAGTAGGAGT TCTCTGCAGAAGGAGAAGGAAATAGCAAAAAGTCTCCAAGAAGCGCGTGCAGCACCGCTCT		

	TACAGGAGCTTAAAAAATTCCACCTCTTCTGTGGTGGAAAAGCCAAACTATTTTTTAAAA AACGCCCTAAAACAATAGTAAAGGCCGGACCAGAATATGGCCAAGGGATGAACCTATTAGCC GCCTGGCGCAAATTCACAGGCCAAAAAGGAAAAGGAGCCGGATTATGTTTTGCTTTCAGAAA GAGGAATGCCTCGACGTCGAGAATTTGTGATGCAGGTGAAGGTAGGCAATGAAGTTGCTACAG GAACAGGACCTAATAAAAAAGATAGCCAAAAAAATGCTGCAGAAGCAATGCTGTTACAACCTG GTTATAAAGCATCCACTAATCTTCAGGATCAACTTGAGAAGACAGGGGAAAACAAAGGATGGA GTGGTCCAAAGCCTGGGTTTCCTGAACCAACAAATAATACTCCAAAAGGAATTCTTCATTGTG CTCCTGATGTTTATCAAGAGATGGAAGCCAGCCGCCACAAAGTAATCTCTGGCACTACTCTAG GCTATTTGTCACCCAAAGATATGAACCAACCTTCAAGCTCTTTCTTCAGTATATCTCCCACAT CGAATAGTTCAGCTACAATTGCCAGGGAACCTCTTATGAATGGAACATCTTCTACAGCTGAAG CCATAGGTTTAAAAGGAAGTTCCTACTCCCCCTTGTCTCCAGTACAACCTTCAAAACAAC TGGAATATTTAGCAAGGATTCAAGGCTTTCAGGCAGCCTTAAGTGCCTTGAAACAATTTCTG AACAAAGGACTGGATCCAATCGATGGAGCGATGAATATCGAAAAGGTTCTCTTGAAAAACAAG CCAAGCATCTGAGAGAGAAAGCGGACAATAACCAGGCACCCCGGGCTCCATCGCTCAGGACT GCAAGAAATCAAACCTCGGCCGTCTAGCAGCTCCCAGAACCCGCGGCTGCCACCGCATCCTTAT AAACCTGTGAGCAGCATGAGGGTGTCTGTGTTGAGGGAATGAATGACTAATACCAAAGGCC GAAATACTG		
	ORF Start: ATG at 192		ORF Stop: TAG at 1788
	SEQ ID NO: 100	532 aa	MW at 58273.9kD
NOV25b, CG114308-02 Protein Sequence	MFSVQLSLGEQTWESESSIKKAQAVANKALTESTLPKPVPKPKSNVNNNPGSITPTVELN GLAMKRGEPAIYRPLDPKPFNYRANYNFRGMYNQRYHCPVKIFYVQLTVGNNEFFGEGKTR QAARHNAAMKALQALQNEPIPERSPQNGESGKMDDDDKDANKSEISLVFEIALKRNMPSFEV IKESGPPHMKSFVTRVSVGEFSAEGEGNSKKLSKKRAATTVLQELKKLPPLPVVEKPKLFFKK RPKTIIVKAGPEYQGGMNPISRLAQIQAKKEKEPDYVLLSERGMPPRRREFVMQVKGVNEVATG TGPNNKIIAKKNAAEAMLLQLGYKASTNLQDQLEKTGENKGWSGPKPGFPEPTNTPKGIHLHS PDVYQEMEASRHKVISGTTGLYLSPKDMNPSSSFFSISPTSNSSATIARELLMNGTSSTA EA IGLKGSSPTPPCSPVQPSKQLEYLARIQGFQAALSALKQFSEQGLDPIDGAMNIEKGSLEKQA KHLREKADNNQAPPGSIAQDCKKNSAV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 25B.

Table 25B. Comparison of NOV25a against NOV25b.		
Protein Sequence	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV25b	39..608 1..532	485/570 (85%) 485/570 (85%)

5

Further analysis of the NOV25a protein yielded the following properties shown in Table 25C.

Table 25C. Protein Sequence Properties NOV25a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV25a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 25D.

Table 25D. Geneseq Results for NOV25a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93660	Human protein sequence SEQ ID NO:13179 - Homo sapiens, 570 aa. [EP1074617-A2, 07-FEB-2001]	1..608 1..570	570/608 (93%) 570/608 (93%)	0.0
AAM40914	Human polypeptide SEQ ID NO 5845 - Homo sapiens, 615 aa. [WO200153312-A1, 26-JUL-2001]	1..548 13..560	537/548 (97%) 542/548 (97%)	0.0
AAM39128	Human polypeptide SEQ ID NO 2273 - Homo sapiens, 479 aa. [WO200153312-A1, 26-JUL-2001]	37..511 5..479	473/475 (99%) 475/475 (99%)	0.0
AAY83023	Human staufer protein - Homo sapiens, 577 aa. [CA2238656-A1, 22-NOV-1999]	39..559 1..555	273/576 (47%) 346/576 (59%)	e-117
AAB94533	Human protein sequence SEQ ID NO:15268 - Homo sapiens, 206 aa. [EP1074617-A2, 07-FEB-2001]	306..511 1..206	206/206 (100%) 206/206 (100%)	e-115

5

In a BLAST search of public sequence databases, the NOV25a protein was found to have homology to the proteins shown in the BLASTP data in Table 25E.

Table 25E. Public BLASTP Results for NOV25a

Protein Accession Number	Protein/Organism/Length	NOV25a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NUL3	CDNA FLJ11290 fis, clone PLACE1009622, weakly similar to maternal effect protein STAUFEN - Homo sapiens (Human), 570 aa.	1..608 1..570	570/608 (93%) 570/608 (93%)	0.0
Q8R4D0	RNA-binding protein staufer splice form A - Rattus norvegicus (Rat), 571 aa.	1..608 1..571	532/609 (87%) 550/609 (89%)	0.0

Q96HM1	Staufen (Drosophila, RNA-binding protein) homolog 2 - Homo sapiens (Human), 538 aa.	37..608 5..538	532/572 (93%) 534/572 (93%)	0.0
Q8R4C9	RNA-binding protein staufen splice form B - Rattus norvegicus (Rat), 512 aa.	1..511 1..512	485/512 (94%) 499/512 (96%)	0.0
Q9UGG6	39k3 protein - Homo sapiens (Human), 479 aa.	37..511 5..479	473/475 (99%) 475/475 (99%)	0.0

PFam analysis predicts that the NOV25a protein contains the domains shown in the Table 25F.

Table 25F. Domain Analysis of NOV25a			
Pfam Domain	NOV25a Match Region	Identities/ Similarities for the Matched Region	Expect Value
dsrm	9..73	26/74 (35%) 53/74 (72%)	2.9e-10
dsrm	118..179	15/74 (20%) 41/74 (55%)	0.006
dsrm	208..272	26/74 (35%) 52/74 (70%)	1.3e-17
dsrm	308..373	27/74 (36%) 56/74 (76%)	1.2e-21
dsrm	496..557	13/74 (18%) 44/74 (59%)	0.06

5

Example 26.

The NOV26 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 26A.

Table 26A. NOV26 Sequence Analysis			
	SEQ ID NO: 101	2218 bp	
NOV26a, CG114349-01 DNA Sequence	CCTCGCGCATCCACTCTCCGGCCGGCCGCTGCCGCGCCTCCTCCGTGCGCCCGCCAGCC TCGCCCGCGCCGTACCATGAGCCAGGCCCTACTCGTCCAGCCAGCGGTGTCTCTACCGCC GCACCTTCGGCGGCGCCCGGGCTTCCCGCTCGGCTCCCGCTGAGCTCGCCCGTGTCCCGC GGGCGGGTTTCGGCTCTAAGGGCTCCTCCAGCTCGGTGACGTCCCGCGTGTACCAGGTGTGCG GCACGTCCGGCGGGCGGGGGCTGGGGTTCGTGCGGGCCAGCCGGCTGGGGACCACCCGCA CGCCCTCCTCCTACGGCGCAGGCGAGCTGCTGGACTTCTACTGGCCGACGCGGTGAACAGG AGTTTCTGACCACGCGCACCAACGAGAAGGTGGAGCTGCAGGAGCTCAATGACCGCTTCGCCA ACTACATCGAGAAGGTGCGCTTCTTGAGCAGCAGAACGCGCTCGCCGCCGAAGTGAACCGGC TCAAGGGCCGCGAGCCGACGCGAGTGGCCGAGCTCTACGAGGAGGAGCTGCGGGAGCTGCGGC GCCAGGTGGAGGTGCTCACTAACGAGCGCGCGCTCGACGTGAGCGCGACAACCTGCTCG ACGACCTGCAGCGCTCAAGGCCAAGCTGCAGGAGGAGATTCAAGTTGAAGGAAGAAGCAGAGA		

	ACAATTTGGCTGCCTTCCGAGCGGACGTGGATGCAGCTACTCTAGCTCGCATTGACCTGGAGC GCAGAATTGAATCTCTCAACGAGGAGATCGCGTTCCTTAAGAAAGTGCATGAAGAGGAGATCC GTGAGTTGCAGGCTCAGCTTCAGGAACAGCAGGTCCAGGTGGAGATGGACATGTCTAAGCCAG ACCTCACTGCCGCCCTCAGGGATATCCGGGCTCAGTATGAGACCATCGCGGCTAAGAACATTT CTGAAGCTGAGGAGTGGTACAAGTCAAGGTGTGACACCTGACCCAGGCAGCCAACAAGAACA ACGACGCCCTGCGCCAGGCCAAGCAGGAGATGATGGAATACCGACACCAGATCCAGTCTCTACA CCTGCGAGATTGACGCCCTCAAGGGCACTAACGATTCCCTGATGAGGCAGATGCGGGAATTGG AGGACCGATTGCCAGTGAGGCCAGTGGCTACCAGGACAACATTGCGCGCTGGAGGAAGAAA TCCGGCACCTCAAGGATGAGATGGCCCCCATCTGCGCGAGTACCAGGACCTGCTCAACGTGA AGATGGCCCTGGATGTGGAGATTGCCACCTACCGGAAGCTGCTGGAGGGAGAGGAGAGCCGGA TCAATCTCCCATCCAGACCTACTCTGCCCTCACTTCCGAGAAACCAGCCCTGAGCAAAGGG GTTCTGAGGTCCATACCAAGAAGACGGTGATGATCAAGACCATCGAGACACGGGATGGGGAGG TCGTCAGTGAGGCGACACAGCAGCAGCATGAAGTGCTCTAAAGACGAGAGACCCCTGCGCCACC AGAGACCTCCTCACCCCTGCTCTCACTGCTCCCTGAAGCCAGCCTTCTTCCATCCCTACAC ACCACCCAGCCTCAGTCTCCCGTCACAGCCTCTGACCCCTCCTCACTGGCCATCCCTCGT GGTCCCAACAGCGACATAGCCCATCCCTGCCTGGTCACAGGCATGCCCCGCCACCTCTGCG GACCCAGCTGTGAGCCTTGGCTGTTGGCAGTGAGTGAGCCTGGCTCTTGTGCTGGATGGAGC CCAGGCGGAGCGGTGGCCCTGTCCCTCCACCTCTGTGACCTGAGGCCTACGCTTTGGCTCT GGAGATAGCCCCAGAGCAGGGTGTGGGATACTGCAGGGCCAGGACTGAGCCCCGAGACCTC CCCAGCCCTAGCCCAGGAGAGAGAAAGCCAGGCAGGTAGCCTGGGGGACTAGCCCTGTGGAG ACTGGGGGGCTTGAAATTGTCCCGTGGTCTCTTACTTCTCTTCCCGAGCCAGGGTGGACT TAGAAAGCAGGGGCTACAAGAGGGAATCCCCGAAGGTGCTGGAGGTGGGAGCAGGAGATTGAG AAGGAGAGAAAGTGGGTGAGATGCTGGAGAAGAGAGAGGAGAGAGGAGCAGAGCGGTCTG AGGCTGTTGGAGGGGCGCCACCTCCCGACGCCCTCCCCCCCCCTGCTGCAGGGGCTCTGGA GAGAAACAATAAA		
	ORF Start: ATG at 81		ORF Stop: TAA at 1488
	SEQ ID NO: 102	469 aa	MW at 53464.1kD
NOV26a, CG114349-01 Protein Sequence	MSQAYSSSRVSSYRRFTFGGAPGFPLGSPVFPFRAGFGSKGSSSVTSRVYQVRSRTSGGA GGLGSLRASRLGTRTPSSYGAGELDFSLADAVNQEFLLTRTNEKVELQELNDRFANYIEKV RFLEQQNALAAEVNRLKGREPTRVAELYEELRELRRQVEVLTNQRARVDVERDNLDDLQRL KAKLQEEIQLKEEAENLAAFRADVDAATLARIDLERRIESLNEEIAFLKKVHEEIERELQAO LQEQQVQVEMDSKPDLTAAALRDIAQYETIAAKNISEAEWYKSKVSDLTQAANKNNDALQ AKQEMMEYRHQIQSYTCEIDALKGTNDSLMRQMLEDRFASEASGYQDNIARLEEIIRHLKD EMARHLREYQDLLNVKMLDVEIATYRKLLEGEESRINLPITYSALNFRETSPEQRGSEVHT KKTVMIKTIETRDGEVVSEATQQQHEVL		
	SEQ ID NO: 103	1865 bp	
NOV26b, CG114349-02 DNA Sequence	CCTCGCGCATCCACTCTCCGGCCGGCCGCTGCCGCGCCTCCTCCGTGCGCCGCCAGCC TCGCCCGCGCCGTCAACATGAGCCAGGCTACTCGTCCAGCCAGCGCTGTCTCTACCGCC GCACCTTCGGCGCGCCCGCGGCTTCCCGCTCGGCTCCCGCTGAGCTCGCCGCTGTTCGCCG GGGCGGGTTTCGGCTCTAAGGCTCTCCAGCTCGGTGACGTCCCGCGTGTACAGGTGTTCGC GCACGTGCGGCGGGCGGGGCTGGGGTCTGCGGCGCAGCCGCTGGGGACCACCCGCA CGCCCTCCTCTACGGCGCAGCGAGCTGCTGACTTCTCACTGCGCCAGCGGTGAACAGG AGTTTCTGACCACGCGACCAACGAGAAGGTGGAGCTGCAGGAGCTCAATGACCGCTTCGCCA ACTACATCGAGAAGGTGCGCTTCTGGAGCAGCAGAACGCGCTCGCCGCGGAAGTGAACCGG TCAAGGGCCGCGAGCCGACGCGAGTGGCCGAGCTTACGAGGAGGAGCTGCGGAGCTGCGGC GCCAGGTGGAGGTGCTACTAACCAGCGCGCGCGCTCGACGTGAGCGCGACAACCTGCTCG ACGACCTGCAGCGGCTCAAGGCCAAGCTGCAGGAGGAGATTCACTGAAGGAAGAAGCAGAGA ACAATTTGGCTGCCTTCCGAGCGGACGTGGATGCAGCTACTCTAGCTCGCATTGACCTGGAGC GCAGAATTGAATCTCTCAACGAGGAGATCGCGTTCCTTAAGAAAGTGCATGAAGAGGAGATCC GTGAGTTGCAGGCTCAGCTTCAGGAATCCGGCACCTCAAGGATGAGATGGCCCCCATCTGC GCGAGTACCAGGACCTGCTCAACGTGAAGATGGCCCTGGATGTGGAGATTGCCACCTACCGGA AGCTGCTGGAGGGAGAGGAGAGCCGATCAATCTCCCATCCAGACCTACTCTGCCCTCAACT TCCGAGAAACCAGCCCTGAGCAAAAGGGTCTGAGGTCCATACCAAGAAGACGGTGATGATCA AGACCATCGAGACACGGGATGGGAGGTCTGAGTGGCCACACAGCAGCAGCATGAAGTGC TCTAAAGACAGAGACCCCTCTGCCACCAGAGACCGTCTCACCCCTGTCTCACTGCTCCCTGA AGCCAGCCTTCTTCCATCCAGGACACACACCCAGCCTCAGTCTCCCTCACAGCCTCTGA CAGGCATGCCCGGCCACCTCTGCGGACCCAGCTGTGAGCCTTGGCTGTGGCAGTGAGTGA GCCTGGCTCTTGTGCTGGATGGAGCCAGGCGGGAGCGGTGGCCCTGTCCCTCCACCTCTGT GACCTGAGGCTACGCTTTGGCTCTGGAGATAGCCCCAGAGCAGGGTGTGGGATACTGCAGG GCCAGGACTGAGCCCCGAGACCTCCCGAGCCCTAGCCAGGAGAGAGAAAGCCAGGCGAGT AGCCTGGGGGACTAGCCCTGTGGAGACTGGGGGGCTTGAAATTGTCCCGTGTCTTACTT TCTTTCCCGAGCCAGGGTGGACTTAGAAAGCAGGGGCTACAAGAGGGAATCCCCGAAGGTG CTGAGGTGGAGCAGGAGATTGAGAAGGAGAGAAAGTGGGTGAGATGCTGGAGAAGAGAGAG		

	GAGGAGAGAGGCAGAGAGCGGTCTGAGGCTGGTGGGAGGGGCGCCACCTCCCCACGCCCTCC CCCCCCTGCTGCAGGGGCTCTGGAGAGAAACAATAAA		
	ORF Start: ATG at 81		ORF Stop: TAA at 1137
	SEQ ID NO: 104	352 aa	MW at 39913.2kD
NOV26b, CG114349-02 Protein Sequence	MSQAYSSSRVSSYRRTFGGAPGFPLGSPVFPFRAGFGSKGSSSVTSRVYQVSRTSGGA GGLGSLRASRLGTTTRTPSSYGAGELLDFSLADAVNQEFLLTRTNEKVELQELNDRFANYIEKV RFLEQQNALAAEVNRLKGREPTRVAELYEEELRELRRQVEVLTNQRRVVDVERNLLDDLQRL KAKLQEEIQLKEEAENNLAAFRADVDAATLARIDLERRIESLNEEIAFLKKVHEEEIRELQAO LQEIIRHLKDEMARHLREYQDLLNVKMLDVEIATYRKLLGEESRINLPQIOTYSALNFRETSP EQRGSEVHTKKTVMIKTIETRDGEVVSEATQQQHEVL		
	SEQ ID NO: 105	1291 bp	
NOV26c, CG114349-03 DNA Sequence	ATGAGCCAGGCCTACTCGTCCAGCCAGCGCTGTCCTCCTACCGCGCACCTTCGGCGGGGCC CCGGGCTTCCCGCTCGGCTCCCGCTGAGCTCGCCCGTGTTCGGCGGGCGGGTTTCGGCTCT AAGGGCTCCTCCAGCTCGGTGACGTCCCGCTGTACCAGGTGTCGCGCACGTCGGGCGGGGCC GGGGGCTTGGGGTCTGCTGCGGGCCAGCCGGCTGGGGACCACCCGACGCGCTCCTCCTACGGC GCAGGCGAGCTGCTGGACTTCTCACTGGCCGACGCGGTGAACCAGGAGTTTCTGACCACGCGC ACCAACGAGAAGGTGGAGCTGCAGGAGCTCAATGACCGCTTCGCCAACTACATCGAGAAGGTG CGCTTCTTGGAGCAGCAGAACGCGGCGCTCGCGCCGAAGTGAACCGGCTCAAGGGCCGCGAG CCGACGCGAGTGGCCGAGCTCTACGAGAAGGAAGAGAACAATTGGCTGCCTTCGAGCGGAC GTGGATGCAGCTACTCTAGCTCGCATTGACCTGGAGCGCAGAATTGAATCTCTCAACGAGGAG ATCGCGTTCCTTAAGAAAGTGCATGAAGAGGAGATCCGTGAGTGCAGGCTCAGCTTCAGGAA CAGCAGGTCCAGGTGGAGATGGACATGTCTAAGCCAGACCTCACTGCCGCCCTCAGGGACATC CGGGCTCAGTATGAGACCATCGCGGCTAAGAACATTTCTGAAGCTGAGGAGTGGTACAAGTCG AAGGTGTCAGACCTGACCCAGGCAGCCAACAAGAACAACGACGCCCTGCGCCAGGCCAAGCAG GAGATGATGGAATACCGACACCAGATCCAGTCTACACCTGCGAGATTGACGCCCTGAAGGGC ACTAACGATTCCCTGATGAGGCAGATGCGGGAATTGGAGGACCGATTTGCCAGTGAGGCCAGT GGCTACCAGGACAACATTGCGCGCTGGAGGAGGAAATCCGGCACCTCAAGGATGAGATGGCC CGCCATCTGCGCGAGTACCAGGACCTGCTCAACGTGAAGATGGCCCTGGATGTGGAGATTGCC ACCTACCGGAAGCTGCTGGAGGGAGAGGAGAGCCGGATCAATCTCCCATCCAGACCTACTCT GCCCTCAACTTCCGAGAAACAGCCCTGAGCAAAGGGGTTCTGAGGTCCATACCAAGAAGACG GTGATGATCAAGACCATGGAGACACGGGATGGGGAGTCTGTCAGTGAGGCCACACAGCAGCAG CATGAAGTGCTCTAAAGACGAGAGACCCTCT		
	ORF Start: ATG at 1		ORF Stop: TAA at 1273
	SEQ ID NO: 106	424 aa	MW at 47999.0kD
NOV26c, CG114349-03 Protein Sequence	MSQAYSSSRVSSYRRTFGGAPGFPLGSPVFPFRAGFGSKGSSSVTSRVYQVSRTSGGA GGLGSLRASRLGTTTRTPSSYGAGELLDFSLADAVNQEFLLTRTNEKVELQELNDRFANYIEKV RFLEQQNALAAEVNRLKGREPTRVAELYEKEENNLAAFRADVDAATLARIDLERRIESLNEE IAFLKKVHEEEIRELQALQEQVQVEMDMSPDLTAALRDIRAQYETIAAKNISEAEWYKS KVDLTQAANKNDALRQAKQEMMEYRHIQSYTCEIDALKGTNDLSMRQMLEDRFASEAS GYQDNIAARLEEEIRHLKDEMARHLREYQDLLNVKMLDVEIATYRKLLGEESRINLPQIOTYS ALNFRETSPQRGSEVHTKKTVMIKTMETRDGEVVSEATQQQHEVL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 26B.

Table 26B. Comparison of NOV26a against NOV26b and NOV26c.		
Protein Sequence	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV26b	77..255 77..255	159/179 (88%) 159/179 (88%)
NOV26c	77..469 77..424	326/394 (82%) 328/394 (82%)

Further analysis of the NOV26a protein yielded the following properties shown in Table 26C.

Table 26C. Protein Sequence Properties NOV26a	
PSort analysis:	0.5102 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.2347 probability located in mitochondrial inner membrane; 0.2347 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV26a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 26D.

Table 26D. Geneseq Results for NOV26a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y92335	Human vimentin - Homo sapiens, 466 aa. [WO200020448-A2, 13-APR-2000]	12..469 9..465	291/464 (62%) 365/464 (77%)	e-157
AAB66348	Human vimentin - Homo sapiens, 466 aa. [EP1067142-A1, 10-JAN-2001]	12..469 9..465	291/464 (62%) 364/464 (77%)	e-156
AA W54351	Vimentin - Homo sapiens, 465 aa. [WO9810291-A1, 12-MAR-1998]	12..469 8..464	291/464 (62%) 364/464 (77%)	e-156
AAU87694	Human pancreatic tumour protein #6 - Homo sapiens, 466 aa. [WO200212331-A2, 14-FEB-2002]	12..469 9..465	290/464 (62%) 363/464 (77%)	e-156
AAB29635	Human pollinosis-associated gene 795-encoded protein, SEQ ID NO:26 - Homo sapiens, 466 aa. [WO200065050-A1, 02-NOV-2000]	12..469 9..465	290/464 (62%) 363/464 (77%)	e-156

- 10 In a BLAST search of public sequence databases, the NOV26a protein was found to have homology to the proteins shown in the BLASTP data in Table 26E.

Table 26E. Public BLASTP Results for NOV26a

Protein Accession Number	Protein/Organism/Length	NOV26a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH32116	Desmin - Homo sapiens (Human), 470 aa.	1..469 1..470	469/470 (99%) 469/470 (99%)	0.0
Q9UHN5	Mutant desmin - Homo sapiens (Human), 470 aa.	1..469 1..470	468/470 (99%) 469/470 (99%)	0.0
AAC39939	Mutant desmin - Homo sapiens (Human), 470 aa.	1..469 1..470	468/470 (99%) 468/470 (99%)	0.0
AAC39938	Mutant desmin - Homo sapiens (Human), 470 aa.	1..469 1..470	468/470 (99%) 468/470 (99%)	0.0
Q8TD99	Mutant desmin - Homo sapiens (Human), 470 aa.	1..469 1..470	468/470 (99%) 468/470 (99%)	0.0

PFam analysis predicts that the NOV26a protein contains the domains shown in the Table 26F.

Table 26F. Domain Analysis of NOV26a

Pfam Domain	NOV26a Match Region	Identities/ Similarities for the Matched Region	Expect Value
filament	107..414	183/359 (51%) 292/359 (81%)	4.8e-171
DUF164	203..421	40/247 (16%) 117/247 (47%)	0.45

5

Example 27.

The NOV27 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 27A.

Table 27A. NOV27 Sequence Analysis

	SEQ ID NO: 107	1903 bp
NOV27a, CG114503-01 DNA Sequence	CGGGGGCGGGGCCGGTAGTGGGAGTGCGGGGCGCGGGTGACAGCGCGGGGTGGCGGCGTGG GACCCAGGGGGCGACAGAGGCAGCAGCAGCCCGAGGCCCTGAGGAGAGGAGACCGGCGGGCGG GCAATGCTGGAGACCTTTCGCGAGCGGCTGCTGAGCGTGCAGCAGGATTTACCTCCGGGCTG AAGACTTTAAGTGACAAGTCAAGAGAAGCAAAAGTGAAAAGCAAACCCAGGTATGAGGATACA TGGGCTGCACTTCACAGAAGAGCCAAAGACTGTGCAAGTGCTGGAGAGCTGGTGGATAGCGAG GTGGTCATGCTTCTGCGCACTGGGAGAAGAAAAAGACAAGCCTCGTGGAGCTGCAAGAGCAG CTCCAGCAGCTCCAGCTTTAATCGCAGACTTAGAATCCATGACAGCAAATCTGACTCATTTA GAGGCGAGTTTGGAGAGGTAGAGAACAACCTGCTGCATCTGGAAGACTTATGTGGGCAGTGT GAATTAGAAAGATGCAACATATGCAGTCCAGCAACTGGAGAATTACAAGAAAAATAAGAGG	

	AAGGAACTTGAAACCTTCAAAGCTGAACTAGATGCAGAGCACGCCAGAAGGTCCTGGAAATG GAGCACACCCAGCAAATGAAGCTGAAGGAGCGGCAGAAGTTTTTGGAGGAAGCCTTCCAGCAG GACATGGAGCAGTACCTGTCCACTGGCTACCTGCAGATGTCAGAGCGGCAGAGCCCATAGGC AGCATGTTCATCCATGGAAGTGAACGTGGACATGCTGGAGCAGATGGTCCTGATGGACATATCG GACCAGGAGGCCCTGGACGTCTTCTGAACTCTGGAGGAGAAGAGAACAACCTTGGTGTCTCCCC GCCTTAGGTAGGGTTGACAACTTGCATTAGCTGAACAGGGCAGTATCGATGCCACTCCCCT CCAAAGGTGAGACGTGAGAACCATCTGCCAGTCACTTACGCATAAAACCCCAAGCTCACAGCC AGCTCCTGGCTCCCTAACCCACGGTTCACACGGCTGTGTGGCAGCTGCAACAGTGGTGTGG TTCCGTGATGAATTCTTCTCAAAGATTGACATGCTCCACTCCGGTAACCTTGGTGAGTTGAG AGCTTTCTTGTGTTTCCCTCCTTACCATCCAGAAATCCATTTGAGTCTGCTCCTTGTGG TTAAGGACTGGCGTTGTCAGGGAGGTGCGGACTCTCTGCGGGGCTCACGGAAACTCTTCCC TCTTCGTGCGACAGGCATTTAGGGGCGTGCTGCCATGGGCAAAGCCATGGTGTGTGTTGAGC TCTTGGCCTGTGTTGTAACCTAGTTGCACTTCAGTTCCTTTTCATCCCTTACAAAAATTTGT TTCACATTCATGCAGCAAATATGGGCTGAGGTGCCAGACCTGTACCTGGGCTTGGTGCCTTTC AAATTTGAGACAGTCTTTGGGCTGGGTCAAGGCAAAGCTCAGTCGTCACAGCAGCACCTCA GCCATCTGTAGAAGGTTCTACCATACCACGGTTTCAGCTTCTCTAAACTTCTACCCGCTT CTCCTGGCAATCTGTGAGAACGGTGTCTCTGGGGAAGAGAAGGAGCTTGGGTGCATTGGCC CTCATCCTGAGAAGGCCAGAATACTGGAGACCAGCGTGAACCTCACCCAGAGTCAGGGGAAG ATTTAGAAACAGTGACACCTGCATATAGAATTTGATTCTTGAAGAGCCTATTAGTTCCAT AAAATTGGAGAACTGCTGAAGGTGAGTAATCCGACTTCTCAGCAGTGGTGTCTCTGAATTA CTGCAAAGGTTAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA TGATGATGTCGAC		
	ORF Start: ATG at 130		ORF Stop: TAA at 988
	SEQ ID NO: 108	286 aa	MW at 32975.1kD
NOV27a, CG114503-01 Protein Sequence	MLETFRERLLSVQDFTSLGLKTLSDKSREAKVKSPPRYEDTWAALHRRKDCASAGELVDSEV VMLSAHWEKKKTSVLQEQQLPALIADLESMTANLTHLEASFEEVENLLHLEDLCGQCE LERCKHMQSQLENYKKNRKELETFKAELDAEHAQKVLMEHTQMKLKERQKFFEEAFQDD MEQYLSTGYLQIAERREPIGSMSSMEVNVDMLEQMLMDISDQEALDVFLNSGGEENTVLSPLA LGRVDKLALAEPPGYRCHSPPKVRRENHLPVITYA		
	SEQ ID NO: 109	966 bp	
NOV27b, CG114503-02 DNA Sequence	CCGAGGCCTGAGGAGAGGAGACCGCGCGCGCGCAATGCTGGAGACCTTCGCGAGCGGCTG CTGAGCGTCAGCAGGATTTACCTCCGGCTGAAGACTTTAAGTGACAAGTCAAGAGAAGCA AAAGTGAAAGCAAAACCCAGGACTGTTCCATTTTTGCCAAAGTACTCTGCTGGATTAGAATTA CTTAGCAGGTATGAGGATACATGGGCTGCACCTTCAGAGAAGGCCAAAGACTGTGCAAGTGCT GGAGAGCTGGTGGATAGCAGGTGGTGTCTTTCTGCGCACTGGGAGAGAAAAGACAAGC CTCGTGGAGCTGCAAGAGCAGCTCCAGCAGCTCCAGCTTTAATCGCAGACTTAGAATCCATG ACAGCAAATCTGACTCATTTAGAGCGAGTTTGGAGGAGGTAGAGAAACCTGCTGCATCTG GAAGACTTATGTGGGCAGTGTGAATTAGAAAGATGCAACATATGCAGTCCCAGCAACTGGAG AATTACAAGAAAAATAAGAGGAAGGAACCTGAAACCTTCAAAGCTGAAGTATAGATGCAGAGCAC GCCAGAGGTCCTGGAAATGGAGCACACCAGCAAAATGAAGCTGAAGGAGCGGCAGAGTTT TTGAGGAAGCCTTCCAGCAGGACATGGAGCAGTACCTGTCCACTGGCTACCTGCAGATTGCA GAGCGGCGAGAGCCCATAGGCAGCATGTATCCATGGAAGTGAACGTGGACATGCTGGAGCAG ATGGACCTGATGGACATATCGGACCAGGAGGCCCTGGACGCTTCTTAACTCTGGAGGAGAA GAGAACACTGTGCTGTCCTCCGCTTAGGTAGGGTTGACAACTTCATTAGCTGTAACCAAGG CAGTATCGATGCCACTCCCTCCAAAGGTGAGACGTGAGAACCATCTGCCAGTCACTTACGCA TAAACCCCAAGCTCACAGCC		
	ORF Start: ATG at 37		ORF Stop: TAA at 946
	SEQ ID NO: 110	303 aa	MW at 34830.2kD
NOV27b, CG114503-02 Protein Sequence	MLETLRERLLSVQDFTSLGLKTLSDKSREAKVKSPPRTVPFLPKYSAGLELLSRYEDTWAALH RRKDCASAGELVDSEVVMLSAHWEKKKTSVLQEQQLPALIADLESMTANLTHLEASFEE VENLLHLEDLCGQCELERCKHMQSQLENYKKNRKELETFKAELDAEHAQKVLMEHTQMKL KERQKFFEEAFQDDMEQYLSTGYLQIAERREPIGSMSSMEVNVDMLEQMLMDISDQEAL DVFLNSGGEENTVLSPLGRVDKLALAEPPGYRCHSPPKVRRENHLPVITYA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 27B.

Table 27B. Comparison of NOV27a against NOV27b.		
Protein Sequence	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV27b	1..286 1..303	273/303 (90%) 273/303 (90%)

Further analysis of the NOV27a protein yielded the following properties shown in Table 27C.

Table 27C. Protein Sequence Properties NOV27a	
PSort analysis:	0.4283 probability located in mitochondrial matrix space; 0.3000 probability located in nucleus; 0.1067 probability located in mitochondrial inner membrane; 0.1067 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV27a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 27D.

Table 27D. Geneseq Results for NOV27a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM79272	Human protein SEQ ID NO 1934 - Homo sapiens, 351 aa. [WO200157190-A2, 09-AUG-2001]	1..254 1..271	252/271 (92%) 252/271 (92%)	e-137
ABG07626	Novel human diagnostic protein #7617 - Homo sapiens, 301 aa. [WO200175067-A2, 11-OCT-2001]	1..236 30..282	233/253 (92%) 234/253 (92%)	e-126
AAB43397	Human ORFX ORF3161 polypeptide sequence SEQ ID NO:6322 - Homo sapiens, 196 aa. [WO200058473-A2, 05-OCT-2000]	154..254 15..116	99/102 (97%) 100/102 (97%)	3e-48
AAM42664	Human kidney related polypeptide SEQ ID NO 533 - Homo sapiens, 101 aa. [WO200155323-A2, 02-AUG-2001]	161..254 6..99	93/94 (98%) 93/94 (98%)	9e-46
AAM99849	Human excretory related polypeptide SEQ ID NO 586 - Homo sapiens, 101 aa. [WO200155313-A2, 02-AUG-2001]	161..254 6..99	93/94 (98%) 93/94 (98%)	9e-46

In a BLAST search of public sequence databases, the NOV27a protein was found to have homology to the proteins shown in the BLASTP data in Table 27E.

Table 27E. Public BLASTP Results for NOV27a				
Protein Accession Number	Protein/Organism/Length	NOV27a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H0U2	Hypothetical 34.8 kDa protein - Homo sapiens (Human), 303 aa.	1..286 1..303	285/303 (94%) 285/303 (94%)	e-158
Q96EV8	Unknown (protein for MGC:20210) (Dysbindin) - Homo sapiens (Human), 351 aa.	1..254 1..271	252/271 (92%) 252/271 (92%)	e-137
Q91WZ8	Dysbindin (Dystrobrevin binding protein 1) - Mus musculus (Mouse), 352 aa.	1..253 1..270	217/270 (80%) 232/270 (85%)	e-118
Q96NV2	CDNA FLJ30031 fis, clone 3NB692001349 - Homo sapiens (Human), 270 aa.	65..254 1..190	189/190 (99%) 189/190 (99%)	e-103
Q9D314	5430437B18Rik protein - Mus musculus (Mouse), 271 aa.	65..253 1..189	158/189 (83%) 172/189 (90%)	3e-86

5

PFam analysis predicts that the NOV27a protein contains the domains shown in the Table 27F.

Table 27F. Domain Analysis of NOV27a			
Pfam Domain	NOV27a Match Region	Identities/ Similarities for the Matched Region	Expect Value

10

Example 28.

The NOV28 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 28A.

Table 28A. NOV28 Sequence Analysis			
	SEQ ID NO: 111	807 bp	
NOV28a, CG114588-01 DNA	CCGAGGCAAGCTCACCTACACCCAGCAGCTGGAGGACCTCAAGAGGCAGCTGGAGGAGGAGGT GGCCCACTGGAGGACCAAGTATGAGACGGACGCCATTGAGCGGACTGAGGAGCTCGAGGAGGC		

190

	AGCTTGAAGAAGAATTGAAACTGTGACGAACAACCTTGAAGTCACTGGAGGCTCAGGCTGAGA AGTACTCGCAGAAGGAAGACAGATATGAGGAAGAGATCAAGGTCTTTCCGACAAGCTGAAGG AGGCTGAGACTCGGGCTGAGTTTGGCGAGAGGTCAGTAACTAAATTGGAGAAAAGCATTGATG ACTTAGAAGACGAGCTGTACGCTCAGAACTGAAGTACAAAGCCATCAGCGAGGAGCTGGACC ACGCTCTCAACGATGACTTCCATATAAGTTTCTTTGCTTCACTTCTCCCAAGACTCCCTCG TCGAGCTGGATGTCCACCTCTCTGAGCTCTGCATTGTCTATTCTCCAGCTGACCCTGGTTC TCTCTCTTAGCATCTCGCCTTAGAGCCAGGCACACACTGTGCTTTCTATTGTACAGAAGCTCT TCGTTTCAGTGTCAAATAAACACTGTGTAAGCTAAAAAAA		
	ORF Start: ATG at 57		ORF Stop: TAA at 909
	SEQ ID NO: 116	284 aa	MW at 32708.1kD
NOV28c, CG114588-03 Protein Sequence	MDAIKKKMQLKLDKENALDRAEQAEADKKAEDRSKQLEDELVSLQKKLKGTEDELDKYSEA LKDAQEKLELAEEKATDAEADVASLNRRILVVEELDRAQERLATALQKLEEAKEADESERG MKVIESRAQKDEEKMEIQEIQLEAKHIAEDADRYEEVARKLVIIESDLERAEEARLESEGK CAELEELKTVTNLKSLEAQAEKYSQKEDRYEEIKVLSDKLKEAETRAEFAERSVTKLEKS IDDLLEDELYAQKLKYKAISELDHALNDMTSI		
	SEQ ID NO: 117	756 bp	
NOV28d, CG114588-04 DNA Sequence	CCGCGCGCTCGCCCCGCGCTCCTGCTGCAGCCCCAGGCCCTCGCCGCGCCACCATGGACG CCATCAAGAAGAAGATGCAGATGCTGAAGCTCGACAAGGAGAAGCGCTTGGATCGAGCTGAGC AGGCGGAGGCCGACAAGAAGGCGCGGAAGACAGGAGCAAGCAGCTGGAAGATGAGCTGGTGT CACTGCAAAAGAACTCAAGGGCACCGAAGATGAAGTGGACAAATACTCTGAGGCTCTCAAAG ATGCCAGGAGAAGCTGGAGCTGGCAGAGAAAAGGCCACCGATGCTGAAGCCGACGTAGCTT CTCTGAACAGACGCATCCAGCTGGTTGAGGAAGAGTTGGATCGTGCCAGGAGCGTCTGGCAA CAGCTTTGCAGAACTGGAGGAAGCTGAGAAGGCAGCAGATGAGAGTGAGAGAGGTCAGTAAC TAAATTGGAGAAAAGCATTGATGACTTAGAAGACGAGCTGTACGCTCAGAACTGAAGTACAA AGCCATCAGCGAGGAGCTGGACCAGCTCTCAACGATATGACTTCCATATAAGTTTCTTTGCT TCACTTCTCCCAAGACTCCCTCGTCGAGCTGGATGTCCCACCTCTCTGAGCTCTGCATTGTG TATTCTCCAGCTGACCCTGGTTCTCTCTCTTAGCATCCTGCCTTAGAGCCAGGCACACACTGT GCTTTCTATTGTACAGAAGCTCTTCGTTTCAGTGTCAAATAAACACTGTGTAAGCTAAAAAAA		
	ORF Start: ATG at 57		ORF Stop: TAA at 438
	SEQ ID NO: 118	127 aa	MW at 14414.9kD
NOV28d, CG114588-04 Protein Sequence	MDAIKKKMQLKLDKENALDRAEQAEADKKAEDRSKQLEDELVSLQKKLKGTEDELDKYSEA LKDAQEKLELAEEKATDAEADVASLNRRILVVEELDRAQERLATALQKLEEAKEADESERG Q		
	SEQ ID NO: 119	1684 bp	
NOV28e, CG114588-05 DNA Sequence	CGTCTCTCGCCCCGACCGCGCTCGCCCCGCGCTCCTGCTGCAGCCCCAGGGCCCTCGCC GCCGCCACCATGGACGCCATCAAGAAGAAGATGCAGATGCTGAAGCTCGACAAGGAGAAGGCC TTGGATCGAGCTGAGCAGGCGGAGGCCGACAAGAAGGCGCGGAAGACAGGAGCAAGCAGCTC GAGGAGGACATCGCGCCAAGGAGAAGTTGCTGCGGGTGTGCGAGGACGAGCGGGACCGGGTG CTGGAGGAGCTGCACAAGGCGGAGGACAGCCTCTGGCCGCCAAGAGGCCGCCGCCAAGGCT GAAGCCGACGTAGCTTCTCTGAACAGACGCATCCAGCTGGTTGAGGAAGAGTTGGATCGTGCC CAGGAGCGTCTGGCAACAGCTTTGCAGAAGCTGGAGGAAGCTGAGAAGGCAGCAGATGAGAGT GAGAGAGGCATGAAAGTCATTGAGAGTCGAGCCCAAAAAGATGAAGAAAAATGGAATTCAG GAGATCCAACCTGAAAGAGGCCAAGCACATTGCTGAAGATGCCGACCGCAATATGAAGAGGTG GCCCGTAAGCTGGTCATCATTGAGAGCGACCTGGAACGTGCAGAGGAGCGGGCTGAGCTCTCA GAAGGCAATGTGCCGAGCTTGAAGAAGAATTGAAAACCTGTGACGAACAACCTGAAGTCACTG GAGGCTCAGGCTGAGAAGTACTCGCAGAAGGAAGACAGATATGAGGAAGAGATCAAGGTCCTT TCCGACAAGCTGAAGGAGGCTGAGACTCGGGCTGAGTTTGCAGGAGAGGTCAGTAACATAAATTG GAGAAAAGCATTGATGACTTAGAAGAGAAAAGTGGCTCATGCCAAAGAAGAAAACCTTAGTATG CATCAGATGCTGGATCAGACTTTACTGGAGTTAAACAACATGTGAAAACCTTCTTAGCTGCCA CCACATTCTTTCTGTTTGTGTTTGTGTTTGTGTTTAAACACCTGCTTACCCCTTAAATGCAATTT ATTTACTTTTACCACTGTACAGAAACATCCACAAGATACCAGCTAGGTCAGGGGGTGGGGAA AACACATACAAAAGGCAAGCCCATGTGAGGGCGATCCTGGTTCAAATGTGCCATTTCCCGGG TTGATGCTGCCACACTTTGTAGAGAGTTTAGCAACACAGTGTGCTTAGTCAGCGTAGGAATCC TCACTAAAGCAGAAGAAGTTCCATTCAAAGTGCCAATGATAGAGTCAACAGGAAGGTTAATGT TGGAACACAATCAGGTGTGGATTGGTGCTACTTTGAACAAAAGGTCCCCCTGTGGTCTTTTG TTCAACATTGTACAATGTAGAATCTGTCCAACACTAATTTATTTGTCTTGAGTTTACTAC AAGATGAGACTATGGATCCCGCATGCCTGAATTCATAAGCCAAAGGCTGTGAAGCCACGCT GCTCTTCCGAGACTTCCATTCTTTCTGATTGGCACACGTGCAGCTCATGACAATCTGTAGGA TAACAATCAGTGTGGATTTCCACTCTTTTCAGTCTTTCATGTTAAAGATTAGACACCACATA CAACTGGTAAAGGACGTTTCTGTAGAGTTTAACTATATGTAACATTGTATAATGATATGG AATAAATGCACATTGTAGGACATTTTCTAAAAAATAAAAAA		

	ORF Start: ATG at 73		ORF Stop: TGA at 925
	SEQ ID NO: 120	284 aa	MW at 32677.1kD
NOV28e, CG114588-05 Protein Sequence	MDAIKKKMQLKLDKENALDRAEQAEADKKAEDRSKQLEEDIAAKEKLLRVSEDERDRVLEE LHKAEDSLAAEEAAKAEADVASLNNRIQLVEEELDRAQERLATALQKLEEAKEKADESERG MKVIESRAQKDEEKMEIQEIQLKEAKHIAEDADRKYEVEARKLVIIESDLERAEEAELSEGK CAELEELKTVTNLKSLEAQAEKYSQKEDRYEEI KVLSDKLKEAETRAEFAERSVTKLEKS IDDLEEKVAHAKEENLSMHQMLDQTLLELNNM		
	SEQ ID NO: 121	1199 bp	
NOV28f, CG114588-06 DNA Sequence	CCCGACCGCGCTCGCCCCGCCCTCCTGCTGCAGCCCCAGGGCCCTCGCGCGCCGCCACCA TGGACGCCATCAAGAAGAAGATGCAGATGCTGAAGCTCGACAAGGAGAACGCCCTTGGATCGAG CTGAGCAGGCGGAGGCCGACAAAGAGCGCGGGAAGACAGGAGCAAGCAGCTCGAGGAGGACA TCGCGGCCAAGGAGAAGTTGCTGCGGGTGTGCGAGGACGAGCGGGACCGGGTGTGAGGAGGC TGCACAAGGCGGAGGACAGCCTCCTGGCGCCGAAGAGGCGCGCCCAAGGCTGAAGCCGACG TAGCTTCTCTGAACAGACGCATCCAGCTGGTTGAGGAAGAGTTGGATCGTCCCAGGAGCGTC TGGCAACAGCTTTGCAGAAGCTGGAGGAAGCTGAGAAGGCAGCAGATGAGAGTGAGAGAGGCA TGAAAGTCATTGAGAGTCGAGCCCAAAAAGATGAAGAAAAATGGAAATTCAGGAGATCCAAC TGAAAGAGGCAAGCACATTGCTGAAGATGCCGACCGCAAATATGAAGAGGTGGCCCGTAAGC TGGTCATCATTGAGAGCGACCTGGAACGTGCAGAGGAGCGGGCTGAGCTCTCAGAAGGCCAAG TCCGACAGCTGGAAGAACAATTAAGAATAATGGATCAGACCTTGAAGCATTAAATGGCTGCAG AGGATAAGTACTCGCAGAAGGAAGACAGATATGAGGAAGAGATCAAGGTCCTTTCCGACAAGC TGAAGGAGGCTGAGACTCGGGCTGAGTTTGCAGGAGAGGTCAGTAACTAAATTGGAGAAAAGCA TTGATGACTTAGAAGAGAAAGTGCTCATGCCAAAGAAGAAAACCTTAGTATGCATCAGATGC TGGATCAGACTTTACTGGAGTTAAACAACATGTGAAAACCTCCTTAGCTGCGACCACATTCTT TCGTTTTGTTTTGTTTTGTTTTTAAACACCTGCTTACCCCTTAAATGCAATTTATTTACTTTT ACCACTGTACAGAAACATCCACAAGATACCACTAGGTCAGGGGGTGGGGAAAACAGATACA AAAAGGCAAGCCCATGTCAGGGCGATCCTGGTTCAAATGTGCCATTTCCCGGGTTGATGCTGC CACACTTGTAGAGAGTTTAGCAACACAGTGTGCTTAGTCAGCGTAGGAATCCTCACTAAAGC AG		
	ORF Start: ATG at 63		ORF Stop: TGA at 915
	SEQ ID NO: 122	284 aa	MW at 32816.4kD
NOV28f, CG114588-06 Protein Sequence	MDAIKKKMQLKLDKENALDRAEQAEADKKAEDRSKQLEEDIAAKEKLLRVSEDERDRVLEE LHKAEDSLAAEEAAKAEADVASLNNRIQLVEEELDRAQERLATALQKLEEAKEKADESERG MKVIESRAQKDEEKMEIQEIQLKEAKHIAEDADRKYEVEARKLVIIESDLERAEEAELSEGQ VRQLEEQLRIMDQTLKALMAEDKYSQKEDRYEEI KVLSDKLKEAETRAEFAERSVTKLEKS IDDLEEKVAHAKEENLSMHQMLDQTLLELNNM		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 28B.

Table 28B. Comparison of NOV28a against NOV28b through NOV28f.		
Protein Sequence	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV28b	1..133 91..223	101/133 (75%) 109/133 (81%)
NOV28c	1..133 127..259	117/133 (87%) 118/133 (87%)
NOV28d	66..133 38..98	23/68 (33%) 37/68 (53%)
NOV28e	1..133 127..259	117/133 (87%) 118/133 (87%)

NOV28f	1..133 127..259	101/133 (75%) 109/133 (81%)
--------	--------------------	--------------------------------

Further analysis of the NOV28a protein yielded the following properties shown in Table 28C.

Table 28C. Protein Sequence Properties NOV28a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV28a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 28D.

Table 28D. Geneseq Results for NOV28a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB90770	Human Tumour Endothelial Marker polypeptide SEQ ID NO 273 - Homo sapiens, 284 aa. [WO200210217-A2, 07-FEB-2002]	1..133 127..259	132/133 (99%) 133/133 (99%)	3e-67
AAG66545	Human interferon-alpha induced polypeptide, TPM1 - Homo sapiens, 284 aa. [WO200159155-A2, 16-AUG-2001]	1..133 127..259	132/133 (99%) 133/133 (99%)	3e-67
AAM78512	Human protein SEQ ID NO 1174 - Homo sapiens, 284 aa. [WO200157190-A2, 09-AUG-2001]	1..133 127..259	132/133 (99%) 133/133 (99%)	3e-67
AA Y92334	Human alpha-tropomyosin - Homo sapiens, 284 aa. [WO200020448-A2, 13-APR-2000]	1..133 127..259	132/133 (99%) 133/133 (99%)	3e-67
ABB57037	Mouse ischaemic condition related protein sequence SEQ ID NO:47 - Mus musculus, 284 aa. [WO200188188-A2, 22-NOV-2001]	1..133 127..259	131/133 (98%) 133/133 (99%)	7e-67

10

In a BLAST search of public sequence databases, the NOV28a protein was found to have homology to the proteins shown in the BLASTP data in Table 28E.

Table 28E. Public BLASTP Results for NOV28a				
Protein Accession Number	Protein/Organism/Length	NOV28a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q91XN6	Tropomyosin alpha isoform - Rattus norvegicus (Rat), 287 aa.	1..161 127..287	156/161 (96%) 158/161 (97%)	1e-81
P18343	Tropomyosin alpha chain, brain-2 (TMBR-2) - Rattus norvegicus (Rat), 251 aa.	1..161 91..251	156/161 (96%) 158/161 (97%)	1e-81
Q9Y427	Hypothetical 34.9 kDa protein - Homo sapiens (Human), 308 aa (fragment).	1..133 151..283	132/133 (99%) 133/133 (99%)	9e-67
P09493	Tropomyosin 1 alpha chain (Alpha-tropomyosin) - Homo sapiens (Human), 284 aa.	1..133 127..259	132/133 (99%) 133/133 (99%)	9e-67
C39816	tropomyosin 5a, fibroblast - rat, 248 aa.	1..133 91..223	131/133 (98%) 133/133 (99%)	2e-66

- 5 Pfam analysis predicts that the NOV28a protein contains the domains shown in the Table 28F.

Table 28F. Domain Analysis of NOV28a			
Pfam Domain	NOV28a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Tropomyosin	1..137	115/137 (84%) 137/137 (100%)	2.5e-109

10 **Example 29.**

The NOV29 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 29A.

Table 29A. NOV29 Sequence Analysis			
	SEQ ID NO: 123	3344 bp	

NOV29a, CG114621-01 DNA Sequence	<p>TTGACTGTATCGCCGAATTCATGAAGTGCCTCTGGTGGCCACTGAGGGCGCAGAGGTCCTC TTCTACTGGACAGATCAGGAGTTTGAAGAGAGTCTCCGGCTGAAGTTCGGGCAGTCAGAGAAT GAGGAAGAAGAGCTCCCTGCCCTGGAGGACCAGCTCAGCACCTCCTAGCCCCGGTCATCATC TCCTCCATGACGATGTGGAGAAGCTCTCGACACCTACACCTGCTTCTCCACGGAAAAATGGC AACTTCCTGTATGTCTTACCTGGTGGCGCCCCCAGACCTGGCGCAGCGTGTCCAGTGTGG GAGCACTTCCAGAGCCTGCTGTGGACCTACAGCCGCTGCGGGAGCAGGAGCAGTGTCTCGCC GTGGAGGCCCTGGAGCGACTGATTACCCCCAGCTCTGTGAGCTGTGCATAGAGGCGCTGGAG CGGCACGTATCCAGGCTGTCAACACCAGCCCCGAGCGGGGAGGCGAGGAGGCCCTGCATGCC TTCCTGCTCGTGCCTCAAGCTGCTGGCATTCTACTTAGCCACAGTGCCAGTCCCTGCGC CCGGCCGACCTGCTTGCCTCATCTCTGTTTCCAGGACCTCTACCCACAGCGAGACACAGCA GAGGACGACATTAGCCTTCCCCGCGAGGGCCCCGAGCAGCCAGAACATCCCCGTGCAGCAG GCCTGGAGCCCTCACTCCAGGGGCCAACTGGGGGGAGCTCTGCAGAGACGGAGACAGACAGC TTCTCCCTCCCTGAGGAGTACTTACACCAGCTCCTTCCCTGGCGATCAGAGCTCAGGTAGC ACCATCTGGCTGGAGGGGGCACCCTCCCATGGATGCCCTCAGATAGCAGGAGACACCTC CAAACACTGGTTCCTCACTGCCCTGTGCCTTCCGGCCCCAGAAGGATCTTCTGGATGCCAAC GTGAAGGAAGCTACTGCCCCCTAGTGCCCCACACCATGTACTGCCTGCCCCCTGTGGCAGGGC ATCAACCTGGTGTCTCTGACCAGGAGCCCCAGCGCGCCCTGGCCCTGGTCTGTCTCCAGCTG ATGGATGGCTTCTCCATGTCTGGAGAAGAAGCTGAAGGAAGGGCGGAGCCCGGGGCTCCCTG CGCTCCAGCCCCCTCGTGGGAGACCTGCGCCAGAGGATGGACAAGTTTGTCAAGAATCGAGGG GCACAGGAGATTAGAGCAGCTGGCTGGAGTTAAGGCCAAGGCTTTCTCCAAAAGTGAGCCC GGATCCTCCTGGGAGCTGTCTCAGGCATGTGGGAAGCTGAAGCGGCAGCTCTGCGCCTTAC CGCTGAACCTTCTGACCACAGCCCCAGCAGGGGAGGCCACACCTGCCCCAGCAGCTGCAG GACCAAGTGACAGAGCTCATGCGGAGAAGCTGACGAGCTGAAGGACTTCTTGTCTGGTGAAG AGCAGGAGGAACATACCATGGTGTCTACCTAGAAGACTTCCAGGCTTGGTGCATCTCATC TATGTGGACCGCACCCTGGGAGATGGTGGCGCCTTCCCTCAACTGCAGTCAAAGACCTCG TCGGAGTTGGGCAAGGGGCGCTGGCTGCCCTTGTCAAACTAAGGTCTGGTCTGTATCCAG CTGGCGCGCAGATACCTGCAGAGGGCTACACCAGCTGCTGTTCCGGAGGGGAGTTCTAC TGCTCCTACTTCTGTGGTTCGAGAATGACATGGGGTACAACTCCAGATGATCGAGGTGCC GTCTCTCCGACGACTCAGTGCCTATCGGCATGTGGGAGGAGACTACTACAGGAAGCTCCTG CGTACTACAGCAAGAACCGCCAACCGAGGCTGTGAGGTGTACGAGCTGTGGCCCTGCAC CTGTCTGTATCCCACTGACCTGCTGGTGCAGCAGGCCGCGCCAGCTGGCCCGGCGCTCTGG GAGGCTCCCGTATCCCCCTGCTCTAGGCCAAGGTGGCCGAGTCTGCCCTTGCATCTGTCTC TCCAGCCACCCTTGTCTGCCACTGTTCCTCATGACGAGAGCTCCTGTCTGCAGTGGCCATCC TGAGGATAGGGCAGAGTGCCAGGGTGGCCCCAGGGCTTCTAAAACCCACCTAGACCACCT CCATGTGAGGTAAGCAAGGCCAGATCCTTCTCTGGAGGAAGAGGGAAGCCAGGGG TCCTGTTTGTAAAACAACGGTGGCAACAGCTCCTCTCCAGAGCTGCTCTGCCTTTATCCTG GGAGATGGGGAGGAAGCCCCATCTCTGCTTCCCTGCGTGGAGGAAGCCCCACAGCAAGCT CTCTCTACCCAGGTAAAGGTGCTCCTTGGCTGGGTTGAATTCAGCGCTGCCACTTCC TCTCTGCACCTCTGGCAAGTTTCTTCTATTTCCCCACGTTTAAAGCGATGGCACCTCCGTCCC AGGGTGGTGTAGGATTACCCAGTGTGGTAGGTGCTCAATAAATGTTGGTCACTGTTATCACT GAAGCCCAACATGCTAGTCTTCTAGACCTTCTGTGAGTCTGATAAGCCCTTGTAAGTCC CAGCCCCCTCATGCTTGGCTGGCGTCTGCCCTAGGGCTGGGGTCTCAAGCCCCCTGGCCCTGG CCCAGAGATTGGATTCCCTTGGCGGCCGTGGAGCCAGGCTTTGATGTCTTCAAAGCTTCT GTGGTGGCCCTGGATTGAGAACCACACCCGAGGGGTACAGCCCTCTCTTCCAAACCGAGAA GTTCTGTCCAGAATGGACCCAGGACCAAGAGACCTGAGAGCCCTGGGACTGGAGTGTCTG CTCCTCTGAGCCAGGAGGCGGTGCTGGGCCAGAGAGGACGGCGTGGCGAAGTACAGCTGCCA CTGCAGCACAGGATCAGATGGCCGTGTGTGTGATGACAGAGCCTCGCCTTCTGTGTCTTTA GTCTTGAGCCAAAATTTGCTCAAAGACTGATCTCTCTTGCAGGGAACAGCTTTGGGGCTG GGGGAAGTAGAACCCACATGTTGGTCTAAACCTGAGAAGGTGGCAGTGAGGAAGTATCCCT CAGGTGACTGGATCTGTGTTCTCTTAAACATCATCTGATGGAATGGCAATGAAAAGCGTGG TTGTGGAATAACAGAAAACATAAAGGAAAAAACTCCAATCCCCTGAGCCCACTGTTCA GGACCCCTGCTTTGTACCTACTATTTCCCTTTAGTTTGTAGCAGCGGCTGGATGTGATATG TCTAGTTAACCAGTCCCCTTGATCTTTCTATATAATAAATAACACAGGAGTGAACATCCTGA ATCAG</p>
	<p>ORF Start: ATG at 22</p> <p>ORF Stop: TAG at 1978</p>
	<p>SEQ ID NO: 124</p> <p>652 aa</p> <p>MW at 73743.7kD</p>
NOV29a, CG114621-01 Protein Sequence	<p>MKCVLVATEGAELVLYWTDQEFEEESRLKFGQSENEEEELPALEDQLSTLLAPV I ISSMTMLE KLSDTYTCFSTENGFLYVLHLVRPPDLAQRVQLWEHFQSLWLYSRLREQEQCFAVEALERL IHPQLCELCIEALERHVIQAVNTSPERGEEALHAFLLVHSLKLLAFYSSHSASSLRPADLLAL ILLVQDLYPSESTAEDDIQPSRRARRSQNI PVQQAWSPHSTGPTGGSSAETETDSFSLPEEY FTPAPSPGDQSSGSTIWLEGGTPPMDALQIAEDTLQTLVPHCPVPSGPRIFLDANVKESYCP LVPHMYCLPLWQGINLVLLTRSPSAPLALVLSQLMDGFSMLEKKLKEGPEPGASLRSQPLVG DLRQRMDKFKVKNRGAQEIQTWLEFKAKAFKSEPGSSWELLQACGKLKRLCAIYRLNFLT APSRGGPHLPQHLQDQVQRLMREKLTWDKDFLLVKSRRNITMVSYLEDDFPLVHFYIVDRTTG QMVAPSLNCSQKTSSELGKGPLAAFVKTKVWSLIQLARRYLQKGYTLLFREGDFYCSYFLWF</p>

	ENDMGYKLQMIIEVPVLSDDSVPIGMLGGDYRKLRLRYYSKNRPTEAVRCYELLALHLSVIPTD LLVQQAGQLARRLWEASRIPLL		
	SEQ ID NO: 125	2109 bp	
NOV29b, CGI14621-02 DNA Sequence	GCCAAGATGAAGTGCCTTGGTGGCCACTGAGGGCGCAGAGGTCTCTTCTACTGGACAGAT CAGGAGTTTGAAGAGAGTCTCCGGCTGAAGTTCGGGCAGTCAGAGAATGAGGAAGAAGAGCTC CCTGCCCTGGAGGACCAGCTCAGCACCTCCTAGCCCCGGTCATCATCTCTCCATGACGATG CTGGAGAAGCTCTCGGACACCTACACCTGCTTCTCCACGGAATAATGGCAACTTCTGTATGTC CTTACCTGTTTGGAGAATGCCTGTTTATTGCCATCAATGGTGACCACCCGAGAGCGAGGGG GACCTGCGGCGGAAGCTGTATGTGCTCAAGTACCTGTTTGAAGTGCACTTTGGGCTGGTGACT GTGGACGGTCATCTTATCCGAAAGGAGCTGCGGCCCCAGACCTGGCGCAGCGTGTCCAGCTG TGGGAGCACTTCCAGAGCCTGCTGTGGACCTACAGCCGCTGCGGGAGCAGGAGCAGTGCTTC GCCGTGGAGGCCCTGGAGCGACTGATTACCCCCAGCTCTGTGAGCTGTGCATAGAGGCGCTG GAGCGGCACGTCATCCAGGCTGTCAACACCAGCCCCGAGCGGGAGGCGAGGAGGCGCTGCAT GCCTTCTGCTCGTGCCTCAAGCTGCTGGCATTCTACTTAGCCACAGTGCCAGCTCCCTG CGCCCGGCGACCTGCTTGCCTCATCTCTGTTTCAAGACCTCTACCCAGCGAGAGCACA GCAGAGGACGACATTAGCCTTCCCCGCGGAGGGCCCCGAGCAGCCAGAATCCCCGTGCAG CAGGCCTGGAGCCCTCACTCCACGGGCCCAACTGGGGGAGCTCTGCAGAGACGGAGACAGAC AGCTTCTCCCTCCCTGAGGAGTACTTCAACACAGCTCCTTCCCTGGCGATCAGAGCTCAGGT AGCACCATCTGGCTGGAGGGGGGACCCCCCATGGATGCCCTTCAGATAGCAGAGGACACC CTCCAAACACTGGTTCCTCCACTGCCTGTGCTTCCGGCCCCAGAAGGATCTTCTGGATGCC AACGTGAAGGAAAGCTACTGCCCTAGTGGCCACACCATGTACTGCTGCCCTGTGGCAG GGCATCAACCTGGTGTCTGACAGGAGCCCCAGCGGCCCTGGCCCTGGTCTGTGCCAG CTGATGGATGGCTTCTCCATGCTGGAGAAGAAGCTGAAGGAAGGGCCGAGCCCGGGGCTCC CTGCGCTCCAGCCCTCGTGGGAGACCTGCGCCAGAGGATGGACAAGTTTGTCAAGAATCGA GGGGCACAGGAGATTGAGACACCTGGCTGGAGTTTAAGGCCAAGGCTTTCTCCAAAGTGAG CCCGGATCCTCTGGGAGCTGCTCCAGGCATGTGGGAAGCTGAAGCGCAGCTCTGCGCCATC TACCGGCTGAACCTTCTGACCACAGCCCCAGCAGGGGAGGCCACACCTGCCCGGGGCTCC CAGGACCAAGTGACAGAGGCTCATGCGGGAGAAGCTGACGGACTGGAAGGACTTCTTGCTGGTG AAGAGCAGGAGGAACATCACCATGGTGTCTACCTAGAAGACTTCCAGGCTTGGTGCACTTC ATCTATGTGGACCGCACCACTGGGAGATGGTGGCGCTTCCCTCACTGCAGTCAAAAGACC TCGTGGAGTTGGGAAGGGGCGCTGGCTGCCTTTGTCAAACTAAGGTCTGGTCTCTGATC CAGCTGGCGCGCAGATACCTGCAGAAGGGCTACACCACGCTGCTGTTCGGGAGGGGATTTC TACTGCTCCTACTTCTGTGGTTCGAGAATGACATGGGGTACAACTCCAGATGATCGAGGTG CCCGTCTCTCCGACGACTCAGTGCCTATCGGCATGCTGGGAGGAGACTACTACAGGAAGCTC CTGCGCTACTACAGCAAGAACCGCCCAACCGAGGCTGTACGGTGTACGAGCTGCTGGCCCTG CACCTGTCTGTATCCCCACTGACCTGCTGGTGCAGCAGGCCGCCAGCTGGCCCGGCGCTC TGGGAGGCTCCCGTATCCCCCTGCTCTAG		
	ORF Start: ATG at 7		ORF Stop: TAG at 2107
	SEQ ID NO: 126	700 aa	MW at 79319.0kD
NOV29b, CGI14621-02 Protein Sequence	MKCVLVATEGAELVFWYTDQEFEESELRKLFQSENEEEELPALEDQLSTLLAPV I ISSMTMLE KLSDTYTCFSTENGFLYVLHFGCELFIAINGDHTSEGLRRLKLYVLKYLFEVHFLVTV GHLIRKELRPPDLAQRVQLWEHFQSLWYSRLREQEQCFAVEALERLIHPQLCELCEALER HVIQAVNTSPERGEEALHAFLLVHSHKLLAFYSSHSASSLRPADLALILLVQDLYPSESTAE DDIQSPRRRASSQNI PVQQAWSPHSTGPTGGSSAETETDSFSLPEEYFTAPSPGDQSSGST IWLEGGTPPMDALQIAEDTLQTLVPHCPVPSGPRRIFLDANVKESYCLPVHPTMYCLPLWQGI NLVLLTRSPSAPLALVLSQLMDGFSMLEKKLKEGPEPGASLRSQLVGDRLRQMDKFVKNRGA QEIQSTWLEFKAKAFKSEPGSSWELLQACGKLKRQLCAIYRLNFLTTPASRGPHLPQHLQD QVQRLMREKLTWKFLLVKSRRNITMVSYLEDFGLVHFIYVDRTTGQMVAPSLNCSQKTSS ELGKGPLAAAFVKTKVWSLIQLARRYLQKGYTTLLFREGDFYCSYFLWFENDMGYKLQMIIEVPV LSDDSVPIGMLGGDYRKLRLRYYSKNRPTEAVRCYELLALHLSVIPTDLLVQQAGQLARRLWE ASRIPLL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 29B.

Table 29B. Comparison of NOV29a against NOV29b.		
Protein Sequence	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV29b	1..652 1..700	617/700 (88%) 618/700 (88%)
--------	------------------	--------------------------------

Further analysis of the NOV29a protein yielded the following properties shown in Table 29C.

Table 29C. Protein Sequence Properties NOV29a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3921 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV29a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 29D.

Table 29D. Geneseq Results for NOV29a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB67493	Drosophila melanogaster polypeptide SEQ ID NO 29271 - Drosophila melanogaster, 596 aa. [WO200171042-A2, 27-SEP-2001]	465..639 409..583	71/187 (37%) 104/187 (54%)	9e-27
AAB93149	Human protein sequence SEQ ID NO:12061 - Homo sapiens, 120 aa. [EP1074617-A2, 07-FEB-2001]	204..280 16..92	22/77 (28%) 34/77 (43%)	0.19
AAM93779	Human polypeptide, SEQ ID NO: 3792 - Homo sapiens, 120 aa. [EP1130094-A2, 05-SEP-2001]	204..280 16..92	22/77 (28%) 34/77 (43%)	0.19
AAU48536	Propionibacterium acnes immunogenic protein #9432 - Propionibacterium acnes, 141 aa. [WO200181581-A2, 01-NOV-2001]	213..302 34..139	37/109 (33%) 41/109 (36%)	0.55
AAE16280	Human kinase PKIN-26 protein - Homo sapiens, 660 aa. [WO200196547-A2, 20-DEC-2001]	212..279 10..77	22/68 (32%) 28/68 (40%)	0.72

10

In a BLAST search of public sequence databases, the NOV29a protein was found to have homology to the proteins shown in the BLASTP data in Table 29E.

Table 29E. Public BLASTP Results for NOV29a				
Protein Accession Number	Protein/Organism/Length	NOV29a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q92902	Hermansky-Pudlak syndrome 1 protein - Homo sapiens (Human), 700 aa.	1..652 1..700	651/700 (93%) 652/700 (93%)	0.0
Q8WXE5	Hermansky-Pudlak syndrome - Homo sapiens (Human), 700 aa.	1..652 1..700	650/700 (92%) 652/700 (92%)	0.0
Q99MK7	Hermansky-Pudlak syndrome protein - Rattus norvegicus (Rat), 706 aa.	1..651 1..705	533/706 (75%) 583/706 (82%)	0.0
O08983	Hermansky-Pudlak syndrome 1 protein homolog - Mus musculus (Mouse), 704 aa.	1..651 1..703	524/704 (74%) 577/704 (81%)	0.0
Q9UH26	DJ1119A7.3 (Putative novel protein similar to HPS (Hermansky-Pudlak syndrome protein)) - Homo sapiens (Human), 149 aa (fragment).	1..121 1..149	107/149 (71%) 113/149 (75%)	5e-53

5

PFam analysis predicts that the NOV29a protein contains the domains shown in the Table 29F.

Table 29F. Domain Analysis of NOV29a			
Pfam Domain	NOV29a Match Region	Identities/ Similarities for the Matched Region	Expect Value

10

Example 30.

The NOV30 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 30A.

Table 30A. NOV30 Sequence Analysis			
	SEQ ID NO: 127	2450 bp	
NOV30a,	CGTCGTCATCTTGGATCCGCGGGACAAGAAAATTCATGCGAGGGAGACGTGGTGGGCGGTCCT		

CG114649-01 DNA Sequence	TCCTGTGACACGACCCTTGAGTGACAGTTCTATTTGATTGCCTCCGGTACTGTGAGGAAAGGA CACGACTCTATGGTGAGGACTGATGGACATACATTATCTGAGAAAAGAACTACCAGGTGACA AACAGCATGTTTGGTGCTTCAAGAAAGAGTTTGTAGAGGGGGTCGACAGTGACTACCATGAC GAAAACATGTACTACAGCCAGTCTTCTATGTTCCACATCGGTCAGAAAAAGATATGCTGGCA TCACCATCTACATCAGGTGAGTCTCAGTTTGGGGCAAGTTTATACGGGCAACAAAACGGA AGTGAAAATGTGACAGGATTGGACCTTTCAGATTTCCAGCATTAGCAGACCGAAACAGGAGG GAAGGAAGTGGTAACCCAACTCCATTAATAAACCCCTTGGCTGGAAGAGCTCCTTATGTTGGA ATGGTAACAAAACCAGCAATGAACAATCCAGGACTTCTCAATACACAATGAAGATTTTCCA GCATTACCAGGCTCCAGCTATAAAGATCCAACATCAAGTAATGATGACAGTAAATCTAATTG AATACATCTGGCAAGACAATTCAGTACAGATGGACCCAAATTCCTGGAGATAAAAGTTCA ACAACACAAAATAATAACCAGCAGAAAAAGGGATCCAGGTGTACCTGATGGTCGGGTACT AACATTCCTCAAGGGATGGTGACGGACCAATTTGGAATGATTGGCCTGTTAACATTTATCAGG GCAGCAGAGACAGACCCAGGAATGGTACATCTGCATTAGGAAGTGACTTAACAACATTAGGC CTCAATCTGAATCTCCTGAAATCTCTACCCCAAATTTGCGTCACCTGGGCATCTTCACCT TGTCGACCTCAAGACATAGACTTCCATGTTCCATCTGAGTACTTAACGAACATTACATTAGG GATAAGCTGGCTGCAATAAACTTGGCCGATATGGTGAAGACCTTCTCTCTATCTCTATTAC ATGAATGGAGGAGACGTATTACAACCTTTAGCTGCAGTGGAGCTTTTAAACCGTATGGAGA TACCACAAGAAGAACGAGTATGGATTACCAGGGCACCAGGCATGGAGCCAACAATGAAAACC AATACCTATGAGAGGGGAACATATTACTTCTTTGACTGTCTTAAGTGGAGGAAGTAGCTAAG GAGTTCCATCTGGAATATGACAAATTAGAAGAACGGCCTCACCTGCCATCCACCTTCAACTAC AACCCTGCTCAGCAAGCCTTCTAAAAAAGAGCTTCCCTTTTCT TGGGGTATGGCTGTCTCAGCACAATACTCAACATAACTGCAGAACTGATGTGGCTCAGGCACC CTGGTTTTAATTCCTTGAGGATCTGGCAATTGGCTTACGCAAAAGGTCAACATTTGAGGTCTCT GCCTTACTAATTTATGTGCTGCCCCAACATAAATTTGTAATTTGTTTTTCTCTAGTTTGAGCA GGGTCTGAATTTTTTTCATTATTTCTTTTTTGCCAGCAGACAGACTTGAGTCTGTAAAGACA AGCAAATACACTGACAGAAGTTTACCATAGTTTCTAAAATGTAAAAAGAAACCCCCAAAAG ACTCAAGAAAATTAGACCACAAATTTGCATTGTTTATTGTAGCACTTGGTAATAAAATAA CAAATGTTGTGCATTTTTATGTGAAGATCCTTCTCGTATTTTCAATTTGGAAGATGAGCAAGA GGTCTGCTTCCTTCATTTTACTTCCCTTCTGTTTTGAAAGGCAGTTTCGCCAAGCTTAATG CAAGAATATCTGACTGTTTAGAAGAAAGATATTGCCACAATCTCTGGATGGTTTCCAGGGTT GTGTTATTACTGAGCTTCATCTTCCAGAAATGAGCAAAACACTGTCCAGTCTTTGTTACGATT TTGTAATAAATGTGTACATTTTTTTAAATTTTTGGACATCACATGAATAAAGGTATGTATGT ACGAATGTGTATATATTATATATATGACATCTATTTTGGAAAATGTTTGGCCCTGCTGTACCTC ATTTTTAGGAGGTGTGATGGATGCAATATATGAAATGGGACATTCTGGAAGTCTGGTCTCAG GGGACTTTGTGCGCCTGTGCACTAAAAGGGCCAGATTTTCCAGCAGCAAGGACATCCATACCC AAGTGAATGTGATGGGACTTAAAAGAAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG CAGCGTTTCATAGGAAGAGAAAAAAGATCAATCTGTATTTCTGACCACATAAAGGCTTCT TCTCTTTGTAATAAAGTAGAAAAGCTCTCCTCAAAAAAAGAGTCTGAG		
	ORF Start: ATG at 136	ORF Stop: TAA at 1345	
	SEQ ID NO: 128	403 aa	MW at 45257.9kD
NOV30a, CG114649-01 Protein Sequence	MVRTDGH TLSEK RNYQV TNSMFGAS RKKFVEGVDS DYHDENMYYSQSSMFPHRSEK DMLASPS TSGQLSQFGASLYGQQNGSENV TGLDLSDFPALADNRNREGSGNPTPLINPLAGRAPYVGMVT KPANEQSQDFSIHNEDFPALPGSSYKDPTSSNDDSKSNLNTSGKTTSSTDGPKFPDGKSSTTQ NNNQKKGIQVLPDGRVTNIPQGMVTDQFGMIGLLTFIRAETDPGMVHLALGSDLTTLGLNL NSPENLYPKFASPWASSPCRQDIDFVHPSEYLTNIHIRDKLAAILKGRYGEDLLFYLMMNG GDVLQLLA AVELFNDRWRYHKEERVWITRAPGMEPTMKTNTYERGTYYFFDCLNWRKVAKEFH LEYDKLEERPHLPSTFNYPNAQQAF		

Further analysis of the NOV30a protein yielded the following properties shown in Table 30B.

Table 30B. Protein Sequence Properties NOV30a	
PSort analysis:	0.7600 probability located in nucleus; 0.2124 probability located in microbody (peroxisome); 0.1589 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV30a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 30C.

5

Table 30C. Geneseq Results for NOV30a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW67820	Human secreted protein encoded by gene 14 clone HE2DE47 - Homo sapiens, 541 aa. [WO9842738-A1, 01-OCT-1998]	66..403 203..540	324/338 (95%) 330/338 (96%)	0.0
AAM79843	Human protein SEQ ID NO 3489 - Homo sapiens, 451 aa. [WO200157190-A2, 09-AUG-2001]	66..403 108..451	323/344 (93%) 329/344 (94%)	0.0
AAW54379	Cell division cycle protein HCDCB - Homo sapiens, 280 aa. [WO9811220-A2, 19-MAR-1998]	124..403 1..280	279/280 (99%) 279/280 (99%)	e-167
AAM78859	Human protein SEQ ID NO 1521 - Homo sapiens, 439 aa. [WO200157190-A2, 09-AUG-2001]	66..294 203..431	215/229 (93%) 221/229 (95%)	e-125
ABB58904	Drosophila melanogaster polypeptide SEQ ID NO 3504 - Drosophila melanogaster, 579 aa. [WO200171042-A2, 27-SEP-2001]	165..395 348..575	120/231 (51%) 166/231 (70%)	2e-66

In a BLAST search of public sequence databases, the NOV30a protein was found to have homology to the proteins shown in the BLASTP data in Table 30D.

Table 30D. Public BLASTP Results for NOV30a				
Protein Accession Number	Protein/Organism/Length	NOV30a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NZN8	Not2p (CCR4-NOT transcription complex, subunit 2) (Similar to CCR4-NOT transcription complex, subunit 2) - Homo sapiens (Human), 540 aa.	66..403 203..540	324/338 (95%) 330/338 (96%)	0.0
Q9H3E0	MSTP046 - Homo sapiens (Human), 365 aa.	66..403 28..365	324/338 (95%) 330/338 (96%)	0.0

Q9D0P1	2600016M12Rik protein - Mus musculus (Mouse), 455 aa.	66..403 118..455	323/338 (95%) 330/338 (97%)	0.0
Q9NWR6	CDNA FLJ20655 fis, clone KAT01590 - Homo sapiens (Human), 490 aa.	66..403 203..490	274/338 (81%) 280/338 (82%)	e-157
Q9P028	HSPC131 - Homo sapiens (Human), 488 aa.	66..337 203..474	256/272 (94%) 263/272 (96%)	e-150

Pfam analysis predicts that the NOV30a protein contains the domains shown in the Table 30E.

Table 30E. Domain Analysis of NOV30a			
Pfam Domain	NOV30a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 31.

The NOV31 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 31A.

Table 31A. NOV31 Sequence Analysis			
	SEQ ID NO: 129	605 bp	
NOV31a, CG116785-01 DNA Sequence	CTGACACCAGCACAGCAAACCCGCGGGATCAAAGTGATACCAGTCGGCAGCATGGCTACGAAA TGTGGGAATTGTGGACCCGGCTACTCCACCCCTCTGGAGGCCATGAAAGGACCCAGGGAAGAG ATCGTCTACCTGCCCTGCATTTACCGAAACACAGGCACTGAGGCCAGATTATCTGGCCACT GTGGATGTTGACCCCAAGTCTCCCCAGTATTGCCAGGTTAGGCGGGGCTTGGGCGCCAGCTAC TTTGAGACCATAGCTGCCCTCATCCCTGGCCCTGGGCCCCCTTCCCAGCTCCATCCTTCTT GGCCCTCCCTGGGGATGCTTGTGCACGCTCAACCTGGGACAAGGGGAGTGCTGAAATCCAGCC TGTGCCGTGCTTCAAACCAAAATGAGTCCACAGGGGCGCCTCTTCAAAGTGGACAGAGGC GTGGCTGGGGAGCACCACCTCTCCCGCATCCTAGGTCATCCACCGGCTGCCCATGCCCAA CCTGAAGGACGAGCTGCATCACTCAGGATGGAACACCTGCAGCAGCTGCTTCGGTGATAGCAC CAAGTCGCGCACCCAGGCTGGTGCTGCCAGTCTCATCTC		
	ORF Start: at 28		ORF Stop: at 343
	SEQ ID NO: 130	105 aa	MW at 11074.6kD
NOV31a, CG116785-01 Protein Sequence	DQSVPVGSMATKCGNCGPGYSTPLEAMKGPREEIVYLPCTYRNTGTEAPDYLATVDVDPKSPQ YCQVRRGLGASYFETIAALIPGPGPPLPSSILLGPPWGCLCT		

10

Further analysis of the NOV31a protein yielded the following properties shown in Table 31B.

Table 31B. Protein Sequence Properties NOV31a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1873 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV31a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 31C.

5

Table 31C. Geneseq Results for NOV31a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE22276	Human selenium- binding protein (HSEBP) #1 - Homo sapiens, 492 aa. [US2002042066-A1, 11-APR-2002]	1..67 13..79	67/67 (100%) 67/67 (100%)	1e-35
AAB57139	Human prostate cancer antigen protein sequence SEQ ID NO:1717 - Homo sapiens, 499 aa. [WO200055174-A1, 21-SEP-2000]	1..67 20..86	67/67 (100%) 67/67 (100%)	1e-35
AAB47946	HSEBP - Homo sapiens, 472 aa. [US6312895-B1, 06-NOV-2001]	9..67 1..59	59/59 (100%) 59/59 (100%)	5e-31
AAE22277	Human selenium- binding protein (HSEBP) #2 - Homo sapiens, 472 aa. [US2002042066-A1, 11-APR-2002]	9..67 1..59	59/59 (100%) 59/59 (100%)	5e-31
AA Y68328	Amyotrophic lateral sclerosis related p53 protein - Homo sapiens, 472 aa. [JP2000000095-A, 07-JAN-2000]	9..67 1..59	59/59 (100%) 59/59 (100%)	5e-31

In a BLAST search of public sequence databases, the NOV31a protein was found to have homology to the proteins shown in the BLASTP data in Table 31D.

10

Table 31D. Public BLASTP Results for NOV31a				
Protein Accession Number	Protein/Organism/Length	NOV31a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q96GX7	Similar to selenium binding protein 1 - Homo sapiens (Human), 472 aa.	9..67 1..59	59/59 (100%) 59/59 (100%)	1e-30
Q13228	Selenium-binding protein 1 - Homo sapiens (Human), 472 aa.	9..67 1..59	59/59 (100%) 59/59 (100%)	1e-30
Q91X87	Unknown (protein for MGC:18519) - Mus musculus (Mouse), 472 aa.	9..77 1..69	58/69 (84%) 59/69 (85%)	4e-28
P17563	Selenium-binding protein 1 (56 kDa selenium-binding protein) (SP56) - Mus musculus (Mouse), 472 aa.	9..77 1..69	58/69 (84%) 59/69 (85%)	4e-28
Q8R1T6	Selenium binding protein 2 - Mus musculus (Mouse), 472 aa.	9..77 1..69	57/69 (82%) 58/69 (83%)	2e-27

Pfam analysis predicts that the NOV31a protein contains the domains shown in the Table 31E.

Table 31E. Domain Analysis of NOV31a			
Pfam Domain	NOV31a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 32.

The NOV32 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 32A.

Table 32A. NOV32 Sequence Analysis			
	SEQ ID NO: 131	2620 bp	
NOV32a, CG118927-01 DNA Sequence	AATTCTTTCATTTTCAGAGTTGAGCAAACTGTGTCTCAAGGAGGTGTGCGCTGCAGGTAGCTGA TGTGGCCAGGATTGAACCCACATTTTATATTCTTGCCATTAGATTCCATTTTCTTATCTC TTATTTCTCAGCGCCAGGTCTCTTTTGTCTTCTGGGCCTCTTCTTCCTTGGCCTCTGCCTC CCTGTCTTCCCATGTTCTTGTATCTGCTGCTGCTGTGTTTTCTCATTGTGTCTCTGCCTAG CTCCCTCTGCCCAGGTTTCTATGTGTCTCTTCTGTGCAGAACTGTCTTCTATTCTTCTCAGT CCCCTTCTCTGCATTTGTTACATCATCTCTTCACTTAATGTATTAGGCTGGGGCTGTTTTGG GACTCCAGAGAGTTACTTTTAACTAGCTACATCTGTCTTCAGGTGCCTGGCTTTCTGGCCTG AGGTGGAAGCAGAGGCCAGCCTCTCTCCGTATCCCTATCTCCCTGGGCTTACATCATGGCG GGGCTTTGAAAATCTCCCTCTGTGGCTTCTTGGAGCCGCATAGATGGGTGATCAGAGCTGGCT CTGGAACCAGGCTGCTCCAGGAGTAAGGAGCCAGTCTTGCATGCAGTGTGAAAAGGTAA TGTCCCTCTTGTGCTGAGCGAGCACCTGGCACACAGCAGGGACCCAGGCAGTGGGGCTGTTAG GTTCTTATCTCTCTGAGCCTTGGGCTTCCGCTATCCTTGGGACGTCTGGTCTTCTGGCTTC CCCGTTCTTCTGCGCCCTGGATGCGGTGACCTGCCAGCACCTGCCGAGCCTTCGTCGGG GAGTCGCCCCATCTCTCCACGCAATCGGCCCTGTGCCCTTGCTGTGCAGCCGGGACCATG TCGACCTCGTCTTGAGGCGCCAGATGAAGAACATCGTCCACAATACTCAGAGGCGGAGATC AAGGTTTCGAGAGGCCACGAGCAATGACCCTGGGGCCATCCAGTCCCTCATGTAGAGATT GCCGACCTCACCTACAACGTTGTGCGCTTCTCGGAGATCATGAGCATGATCTGGAAGCGGCTC AATGACCATGGCAAGAACTGGCGTACGTTTACAAGGCCATGACGCTGATGGAGTACCTCATC AAGACCGGCTCGGAGCGGTGTGCGAGCAGTGAAGGAGAACATGTACGCCGTGCAGACGCTG AAGGACTTCCAGTACGTGGACCGGACGGCAAGGACCAGGGCGTGAACGTGCGTGAGAAAGCT		

	<p>AAGCAGCTGGTGGCCCTGCTGCGCAGCAGGACCGGCTGCGGGAAGAGCGGGCGCACGCGCTC AAGACCAAGGAAAAGCTGGCACAGACCGCCACGGCCTCATCAGCAGCTGTGGGCTCAGGCCCC CCTCCCGAGGCGGAGCAGGCGTGGCCGAGAGCAGCGGGAGGAGGAGCTGCAGCTCCAGCTG GCCCTGGCCATGAGCAAGGAGGAGCGCCAGCAGGAGGAGCGGATCCGTCGCGGGGATGACCTG CGGCTGCAGATGGCAATCGAGGAGAGCAAGAGGGAGACTGGGGGCAAGGAGAGTCTGTCCTC ATGGACCTTGCTGACGTCTTACGGCCCCAGCTCCTGCCCGACACAGACCCCTGGGGGGG CCAGCACCCATGGCTGCTGCGCTCCCGACGGCTGCCCCACCTCGGACCCCTGGGGCGGCCCC CCTGTCCCTCCAGCTGCTGATCCCTGGGGAGGTCCAGCCCCACGCCGGCTCTGGGGACCCC TGGAGGCCTGCTGCCCCGAGGACCCCTCAGTTGACCTTGGGGTGGGACCCAGCCCTGCA GCTGGGGAGGGGCCACGCTGATCCATGGGGAAGTCCGATGGTGGGGTCCCGGTCAGTGGG CCCTCAGCCTCCGATCCCTGGACACCGGCCCGGCCTTCTCAGATCCCTGGGGAGGGTCACCT GCCAAGCCAGCACCAATGGCACAACAGCAGCCGGGGGATTGACACAGGAGCCCGACAGATT TCTGACTTTGACCGACTCCGACGGCACTGCCGACCTCCGGGAGCAGCGCAGGAGAGCTGGAG CTGCTGGCAGGAGAGGTGCGGCCCGAAGCCCTGGGGCGTTGACATGAGTGGGGTCAGGGGA TCTCTGGCTGAGGCTGTGGGCAGCCCCACCTGCAGCCACCAACTCCCAGCCCCCCCCACC CGGAAGACGCGGAGTCAATCTGGGGCCCAATGCAGCCCTCGTCGACCTGGACTCGTGTGGTG AGCCGGCGGGGCCACGCCGCTGGAGCCAAGGCTCCAACCCCTTCTGCCAGGCGGAGGC CCAGCCACTGGCCCTTCCGTCACCAACCCCTTCCAGCCCGGCCTCCCGCAGCGCTCACCCTG AACCAGCTCCGTCTCAGTCTGTGCTCCGCTCCGTCGAGCGCCACCCAGTACATCTCTCCC CTTGGCGGGGCCCTGGCCTGCCCCCATGATGCCCCGGGCCCGGCCCGCCCAACTAAT CCCTTCTCTATAATCCAGGGCGGAAGGGGGCTGGCTCCATCCGCTGCCCCATTCCGGCT CCCTGGGAGATCAGTGTGTGAGTGCATGTGAATGG</p>
	<p>ORF Start: ATG at 880</p> <p>ORF Stop: TAA at 2533</p>
	<p>SEQ ID NO: 132</p> <p>551 aa</p> <p>MW at 57574.6kD</p>
NOV32a, CG118927-01 Protein Sequence	<p>MSTSSLRQMKNIVHNYSEAEIKVREATSNDPWGPSSSLMSEIADLTYNVVFSEIMSMIWK LNDHGKNWRHVYKAMTMEYLIKTGSEVSQCKENMYAVQTLKDFQYVDRDGKQGVNVREK AKQLVALLRDEDRLEERAHALKTEKLAQTATASSAVGSGPPPEAEQAWPQSSGEEELQLQ LALAMSKEEADQEERIRRGDDLRLQMAIEESKRETGGKESSLMDLADVFTAPAPAPPTD PWG GPAPMAAAVPTAAPTSDPWGPPVPPAADPWGPPAPTASGDPWRPAAPAGPSVDPWGGTPAP AAGEGPTPDPWGSDDGVPVSGPSASDPWTPAPAFSDPWGSPAKPSTNGTTAAGGFDEPDE PSDFDLRLTALPTSGSSAGELELLAGEVPARSPGAFDMSGVGRSLAEAVGSPPPAATPTPTPP TRKTPESFLGPNALVDLSVSRPGPTPPGAKASNPLPGGPGPATGPSVNTNPFQAPAPATLT LNQLRLSPVPPVPGAPPTYISPLGGGPGLPMPMPGPPAPNTNPFLL</p>
	<p>SEQ ID NO: 133</p> <p>2449 bp</p>
NOV32b, CG118927-02 DNA Sequence	<p>AATTCTTTCATTTCAGAGTTGAGCAAACTGTGTCTCAAGGAGGTGTGCGTGCAGGTAGCTGA TGTGGCCAGGATTGAACCCACATTTTATATTCTTTGCCATTAGATTCCATTTTCTTATCTC TTATTTCTCAGCGCCAGGTCTCTTTTGTCTTGGGCTCTTCTCTCCTTGGCCTCTGCGCTC CCTGTCTTCCCATGTTCTGTATCTGCTGCTGCTGTGTTTTCTCATTGTGTCTCTGCTAG CTCCCTCTGCCAGGTTCTATGTGTCTCTTCTGTGCAAGTGTCTTCTTCTTCTCTCAGT CCCCCTCTCTGCATTGTGTACATCATCTCTCATCTTAATGTATTAGGCTGGGGCTGTTTTGG GACTCCAGAGAGTTACTTTTAACTAGCTACATCTGTCTTCAGGTGCCTGGCTTTCTGGCCTG AGGTGGAAGCAGAGGCCAGCCTCTCTCCGTATCCCTATCTCCCTGGGCTTACATCATGGCG GGGCTTTGAAATCTCCCTCTGTGGCTTCTTGGAGCCGATAGATGGGTGATCAGAGCTGGCT CTGGAACCAAGGCTGCTCAGGAGTAAGGAGCCAGTCTTGCATGAGTGTGAAAGGTAA TGTCCCTCTTGTGCTGAGCGAGCACCTGGCACACAGCAGGACCCAGGCAGTGGGGCTGTAG GTTCCCTATCTCTCTGAGCCTTGGGCTTCCGCTATCCTTGGGACGTCTGGTCTTCTGGCTTC CCCGGTTCTTCTGCGCCCTGGATGCGGTGACCTGCCAGCACCTGCCGAGCCTTCGTCGCG GAGTCGCCCCATCTCTCCAGCAATCGGCCCTGTGCCCTTGTGCTGCAGCCGGGACCATG TCGACCTCGTCTTGGGCGCCAGATGAAGAACATCGTCCACAACCTACTCAGAGGCGGAGATC AAGGTTGAGAGGCCACGAGCAATGACCCCTGGGGCCATCCAGCTCCCTCATGTGAGAGATT GCCGACCTCACCTACAACGTTGTGCGCTTCTCGGAGATCATGAGCATGATCTGGAAGCGGCTC AATGACCATGGCAAGAACTGGCGTCACGTTTACAAGGCCATGACGCTGATGGAGTACCTCATC AAGACCGGCTCGGAGCGCGTGTGCGAGCAGTGCAAGGAGAACATGTACGCCGTGCAGACGCTG AAGGACTTCCAGTACGTGGACCGCGACGGCAAGGACAGGGCGTGAACGTGCGTGAGAAAGCT AAGCAGCTGGTGGCCCTGCTGCTGGCCATGAGCAAGGAGGAGGCCAGCAGGAGAGCGGATC CGTCGCGGGGATGACCTGCGGCTGCAGATGGCAATCGAGGAGAGCAAGAGGGAGACTGGGGG AAGGAGGAGTCTCCCTCATGACCTTGTGACGTCTTACGGCCCCAGCTTCCGCCCGGCTTCTCAGAT ACAGACCCCTGGGGGGGCCAGCACCCATGGCTGCTGCGCTCCCCACGGCTGCCCCACCTCG GACCCCTGGGGCGGCCCCCTGTCCCTCCAGCTGCTGATCCCTGGGGAGGTCCAGCCCCACG CCGGCTCTGGGGACCCCTGGAGGCTGTGCCCCCTGCAGGACCTCAGTTGACCCCTGGGGT GGGACCCAGCCCCCTGCAGCTGGGAGGGGCCACGCTGATCCATGGGGAAGTCCGATGGT GGGGTCCCGGTACGTGGGCCCTCAGCCTCCGATCCCTGGACACCGGCCCGGCTTCTCAGAT CCCTGGGGAGGGTCACTGCCAAGCCAGCACCAATGGCACAACAGCAGCCGGGGGATTGAC ACGGAGCCGACGAGTCTCTGACTTTGACCGACTCCGCACGGCACTGCCGACCTCCGGGAGC</p>

	AGCGCAGGAGAGCTGGAGCTGCTGGCAGGAGAGGTGCCGGCCCCAAGCCCTGGGGCGTTTGAC ATGAGTGGGGTCAGGGGATCTCTGGCTGAGGCTGTGGGCAGCCCCACCTGCAGCCACACCA ACTCCACGCCCCCACC CGAAGACGCCGGAGTCATTCTGGGGCCAATGCAGCCCTCGTC GACCTGGACTCGCTGGTGAGCCGGCCGGGCCCCACGCCGCTGGAGCCAAGGCCTCCAACCCC TTCTGCCAGGCGGAGGCCAGCCACTGGCCCTTCGTCACCAACCCCTCCAGCCCGCGCCT CCCGCGAGCGTCAACCCTGAACCAGCTCCGTCTCAGTCTGTGCCTCCCGTCCCTGGAGCGCCA CCCACGTACATCTCTCCCTTGGCGGGGGCCCTGGCCTGCCCCCATGATGCCCCGGGGCCCC CCGGCCCCCAACACTAATCCCTTCCTCCTATAATCCAGGGCGGAAGGGGGCCTGGCTCCATCC GGCTGCCCCATTCCGGCTCCCTGGGAGATCAGTGTGTGAGTGCATGTGAAATGG		
	ORF Start: ATG at 880		ORF Stop: TAA at 2362
	SEQ ID NO: 134	494 aa	MW at 51422.0kD
NOV32b, CG118927-02 Protein Sequence	MSTSSLRRQMKNIVHNYSEAEIKVREATSNDPWGPSSSLMSEIADLTYNVVFSEIMSMIWK LNDHGKNWRHVYKAMTLMEYLIKTSERSVSQCKENMYAVQTLKDFQYVDRDQGVNVREK AKQLVALLAMSKEEADQEERIRRGDDLRLQMAIEESKRETGGKESSLMDLADVFTAPAPAP TTDPWGGPAPMAAAVPTAAPTSDPWGPPVPPAADPWGGPAPTASGDPWRPAPAGPSVDPW GGTPAPAAGEGPTPDWGSDDGVPVSGPSASDPWPAPAFSDPWGGSFAKPSNNGTTAAGGF DTEPDEFSDFDRLRTALPTSGSSAGELELLAGEVPARSPGAFDMSGVRGSLAEAVGSPPPAAT PTPTPPTRKTPESFLGPNAALVDLSLVSRLPGPTPPGAKASNPFLPGGGPATGPSVTNPFQPA PPATLTNLQRLRLSPVPPVPGAPPTYISPLGGGPGLPMPMPGPPAPNTNPFLL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 32B.

Table 32B. Comparison of NOV32a against NOV32b.		
Protein Sequence	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV32b	1..551 1..494	360/551 (65%) 362/551 (65%)

5

Further analysis of the NOV32a protein yielded the following properties shown in Table 32C.

Table 32C. Protein Sequence Properties NOV32a	
PSort analysis:	0.4600 probability located in mitochondrial matrix space; 0.4500 probability located in cytoplasm; 0.1903 probability located in lysosome (lumen); 0.1562 probability located in mitochondrial inner membrane
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV32a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 32D.

Table 32D. Geneseq Results for NOV32a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB42049	Human ORFX ORF1813 polypeptide sequence SEQ ID NO:3626 - Homo sapiens, 551 aa. [WO200058473-A2, 05-OCT-2000]	1..551 1..551	550/551 (99%) 550/551 (99%)	0.0
AAB95100	Human protein sequence SEQ ID NO:17064 - Homo sapiens, 576 aa. [EP1074617-A2, 07-FEB-2001]	1..551 1..576	551/576 (95%) 551/576 (95%)	0.0
AAB24234	Human vesicle associated protein 13 SEQ ID NO:13 - Homo sapiens, 576 aa. [WO200060082-A2, 12-OCT-2000]	1..551 1..576	551/576 (95%) 551/576 (95%)	0.0
AAB93525	Human protein sequence SEQ ID NO:12872 - Homo sapiens, 584 aa. [EP1074617-A2, 07-FEB-2001]	1..551 1..584	305/636 (47%) 367/636 (56%)	e-139
ABG12620	Novel human diagnostic protein #12611 - Homo sapiens, 417 aa. [WO200175067-A2, 11-OCT-2001]	2..348 4..345	245/355 (69%) 253/355 (71%)	e-121

In a BLAST search of public sequence databases, the NOV32a protein was found to have homology to the proteins shown in the BLASTP data in Table 32E.

Table 32E. Public BLASTP Results for NOV32a

Protein Accession Number	Protein/Organism/Length	NOV32a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y6I3	EH domain-binding mitotic phosphoprotein - Homo sapiens (Human), 551 aa.	1..551 1..551	551/551 (100%) 551/551 (100%)	0.0
Q9HA18	CDNA FLJ12392 fis, clone MAMMA1002699, highly similar to Rattus norvegicus EH domain binding protein Epsin mRNA - Homo sapiens (Human), 576 aa.	1..551 1..576	551/576 (95%) 551/576 (95%)	0.0
O88339	EH domain binding protein Epsin - Rattus norvegicus (Rat), 575 aa.	1..551 1..575	521/576 (90%) 527/576 (91%)	0.0
O13027	Mitotic phosphoprotein 90 - Xenopus laevis (African clawed frog), 609 aa.	10..551 1..609	366/658 (55%) 407/658 (61%)	e-170

O95207	Epsin 2a - Homo sapiens (Human), 584 aa.	1..551 1..584	304/636 (47%) 366/636 (56%)	e-138
--------	--	------------------	--------------------------------	-------

PFam analysis predicts that the NOV32a protein contains the domains shown in the Table 32F.

Table 32F. Domain Analysis of NOV32a			
Pfam Domain	NOV32a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ENTH	17..140	69/131 (53%) 118/131 (90%)	1.4e-67
UIM	182..199	12/18 (67%) 16/18 (89%)	0.014

5

Example 33.

The NOV33 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 33A.

Table 33A. NOV33 Sequence Analysis			
	SEQ ID NO: 135	806 bp	
NOV33a, CG118981-01 DNA Sequence	ATGACATGGCGGGCAGTGGCTGGCCGGGCCTCCTCATGTCCTGCCTGAAGGGTCCCCATGTCA TCCTCAAGATGGAGGCCATGAAGATTGTCCACCCTGAAAAGTTCCCTGAGCTACCGGCTGCCC CTGCTTCCCGCTGCTCCCGGCCACCCCAACTCTGGCACCCAAGCGTGCCTGGCCCTCAG ACACAGAGATCATGTCAACCAGGCTGTGGGGGGACATGCCTGCCTTGGGAAGGGCACCCC ATACCCCGCCACTGCCACGGCGGCCCGTAAGGGAAGCTCGGAGCTGGGCTTCCCGCGTGG CCCCAGAGGATGAGGTCAATTGTGAATCAGTACGTATTTCGGCCTGGCCCCCTCGGCCTCGGCGG CTTCTTCGGCGGGCAGGCGAGCCCTGGAGTGCCCCACCTGTGGGCACTCCTACAATGTCA CCCAGCGAGGCCCGCGTGTCTCTGCCTGCACTCTGTGTGTGAGCAGTGCCTGCAGATTC TCTACGAGTCTGCCCAAGTACAAGTTTCTCTGCCCCACCTGCCGCGTGAGACTGTGC TCTTACCGACTACGGCTGGCCGCGCTGGCTGTCAACAGTCCATCCTGAGCCGCTGCCGC CTGAGGCGCTGACGGCCCATCCGGGGTCACTGGGGGGCTGAGCCCGAGGCGAGTGTACC AGACCTTCCGGCAGTACTGTGGGGCCGCTGCACCTGCCACGTGCGGAACCCACTGTCCGCCT GCTCCATCATGTAGTAGCGCTGCCTGCCCGCCACTGCCCGCCATGTCTAT		
	ORF Start: ATG at 6		ORF Stop: TAG at 768
	SEQ ID NO: 136	254 aa	MW at 27284.2kD
NOV33a, CG118981-01 Protein Sequence	MAGSGWPGLLMSCLKGPHVILKMEAMKIVHPEKFPELPAAPCFPPAPRPTPLAPKRAWPSDT EIIVNQACGGDMPALEGAPHTPLPRRPRKGSSELGFPRVAPEDVIVNQVIRPGPSASAAS SAAAGEPLECPTCGHSYNVTQRRPRVLSCLHSVCEQCLQILYESCPKYKFISCPTRRET VLF TDYGLAALAVNTSILSRLPPEALTAPSGGQWGAEPGSCYQTFRQYCGAACTCHVRNPLSACS IM		

10

Further analysis of the NOV33a protein yielded the following properties shown in Table 33B.

Table 33B. Protein Sequence Properties NOV33a	
PSort analysis:	0.8950 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV33a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 33C.

5

Table 33C. Geneseq Results for NOV33a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU16313	Human novel secreted protein, Seq ID 1266 - Homo sapiens, 225 aa. [WO200155322-A2, 02-AUG-2001]	30..254 1..225	221/225 (98%) 221/225 (98%)	e-134
ABB05662	Human signal transduction protein clone amy2_10h17 - Homo sapiens, 180 aa. [WO200198454-A2, 27-DEC-2001]	75..254 1..180	180/180 (100%) 180/180 (100%)	e-106
ABB91088	Herbicide active polypeptide SEQ ID NO 299 - Arabidopsis thaliana, 1579 aa. [WO200210210-A2, 07-FEB-2002]	132..208 2..75	31/77 (40%) 38/77 (49%)	3e-06
AAU27735	Mouse full-length polypeptide sequence #60 - Mus musculus, 771 aa. [WO200164834-A2, 07-SEP-2001]	134..239 48..163	39/124 (31%) 59/124 (47%)	1e-04
ABG21275	Novel human diagnostic protein #21266 - Homo sapiens, 774 aa. [WO200175067-A2, 11-OCT-2001]	134..239 44..159	39/124 (31%) 59/124 (47%)	1e-04

In a BLAST search of public sequence databases, the NOV33a protein was found to have homology to the proteins shown in the BLASTP data in Table 33D.

Table 33D. Public BLASTP Results for NOV33a				
Protein Accession Number	Protein/Organism/Length	NOV33a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q9H0X6	Hypothetical 19.4 kDa protein - Homo sapiens (Human), 180 aa.	75..254 1..180	180/180 (100%) 180/180 (100%)	e-106
AAH30073	Similar to hypothetical protein DKFZp761H1710 - Mus musculus (Mouse), 183 aa.	75..254 1..183	169/183 (92%) 172/183 (93%)	3e-98
Q8QZS5	Similar to unknown (Protein for MGC:4734) - Mus musculus (Mouse), 190 aa.	135..207 12..84	28/73 (38%) 41/73 (55%)	5e-08
Q96D59	Hypothetical 21.7 kDa protein - Homo sapiens (Human), 192 aa.	135..207 12..84	27/73 (36%) 41/73 (55%)	1e-07
BAC03481	CDNA FLJ33257 fis, clone ASTRO2005593 - Homo sapiens (Human), 247 aa.	120..207 5..91	36/89 (40%) 49/89 (54%)	2e-07

Pfam analysis predicts that the NOV33a protein contains the domains shown in the Table 33E.

Table 33E. Domain Analysis of NOV33a			
Pfam Domain	NOV33a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C3HC4	136..182	15/55 (27%) 32/55 (58%)	0.025

5

Example 34.

The NOV34 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 34A.

Table 34A. NOV34 Sequence Analysis			
	SEQ ID NO: 137	1955 bp	
NOV34a, CG119385-01 DNA Sequence	GGAAGGTGTGAGGTTGCCGCCATGCCTGGCAGAACGGAGGGAGGCAGTTGGCTCCGGAATGCG GCCGCCGAGATGTTCTCCGCAACCTTCCGGAAGTGGAATGGCGGGAGCCTCAGCATTGCTGC CCACCGACCCCCCGGAAGCGGAACAGAATCCCCGCGTGCCCTTCTCTACTACCTTCCAAAT CCCGCTGCAGCATTGCCGACAGACGATGCCGAAACGAAAGAAGCAGAATCATCACCAGCCA CCGACACAGCAGCAGCCCCCGCTGCCCGAGCGGGAAGAGACTGGAGATGAGGAGGATGGGAGT CCCATCGCTCTCACAGAGGTCCTCAGGATCAAGGGGACCACTGATTCCACCACCTGCTGAGT CTCCCACTCTCTCTTGGGGTAGAGGCCCAATTCGGAGAGGGCTTGGCCCCAGGTCTAGCCCA TATGGTCGTGGTTGGTGGGGAGTCAATGCAGAACCTCCTTTTCCGGGGCCAGGCCATGGGGGT CCCACCAGGGGAAGCTTTCACAAGGAACAGAGAAACCTCGAAGGCTCAAAAGCTGGTCTCTT ATCAAGAATACCTGCCCGCCCAAGGATGACCCCAAGTTATGGAAGACAAATCCGACCGCCCT GTCTGCCGACATTTTGCCAAAAGGGCCACTGTGATATGAGGACCTCTGTGCCTTCTACCAT CCAGGCGTCAATGGACCTCTCTGTGAGACTGTGCCTTCCCATCCAGGCTGGAAGGAGCTCTC TGIGACTAGCGGCCATTATTCTCTGTAGCCCTATGATGGCTACTGTGAGGCTCTTCTAAC ACCCTCAGTCAGTGACACACCCATCCCATCCACCACTTCCCCCGTGTGGGGTCCAGAGTGGTG TTGCATCACTGGTGCGCGGCATACGCGCTTCTTCTGATCCAGCCTGTAGAGACTCGCCTTTG GGACCATCTTTGCTTCTTTCAGTTGCCTCCTGGATCTCTTTCCCGTCATCAATGACTGC		

	TGAACAGGAAACCTCTTTGGTGCTGTTTCTTGTGCATCTGTCCACCTGTTCCCCAGTATTGCC CTCAATTCTTGAGAGCCCTGGAGCGGTTTCTACCATTCCCTTCTTTTAGCTGCTTGTTTTAA GTCCTTTTATGTGACATTCCCTACCCCAATGTTGTGAGCTGCTTGTGAAACTCAGCCAGGT TGTCTAACCTGGGGTCAAGTTTGGGTGACTGGTGCAGAGTACTTCTTAAAGGCCACTCTCC CTGCCTTTGGATTTCATAGTTTCTCTGTCAGTAGCATGATCCCCACCGCTATGGTCTATCTAT GATCACCGTGCTTTGTGAAACTGTGCATCCCTTGTAGCCTTTCTCAGTGTCCGTGGCATTTT TGTGACTTCCAGCACTAGAATAAGTTTCTGCCAAAATGAGTGAGGCGCTTGGTGCCCTCT GGACTTTCCCACTTCCCAACATGGGAGAATTGTGAACTTTCCATCAGACTGCCTCCCTGGCCC TCCCCATTCTTCTCTGTTGGTTATTCTGAGTCTGACACAGACCCATGACATGTCTTATAAAG CCTCCAATGGCTTTATCCTACCTAGATCCCTTCCAGCCCATTTTAATTAGACTATGTCAATTGT GAGGCCACCACTCCATTCAATTGAAATCTGTGAATCTCCACCTTGCCCTATCTTTGGGTAGAAG CTGGACAGTACTGTTGCCCTCTTCCAATCCTCTTCCCCTACATCCCTGGCACTGGTTGTTTTC TGTGAAACAGCAGTGAACAGGTTTCAAGTTTGAACGGCCCTGAGGAAATGGGTGAGGAGTTG TATTGGCAAGAGGGAGGGGTGAGAGCTGTTGGAGAACTGAGAATGAGGTTTTTTTTTTTTTTT TCTTTTAACTTTTTTATATTAGTAATAAATGCAGTGGAAACCAGCATTTTATTATAAAAAA AA		
	ORF Start: ATG at 22		ORF Stop: TGA at 718
	SEQ ID NO: 138	232 aa	MW at 25830.1kD
NOV34a, CG119385-01 Protein Sequence	MPGRTEGGSWLRNAAAADVLRNLPEVEWREPQHCCPPTPRKRKQNPVPLPHYPPNPAAAI DTMPKRKKQNHQPPPTQQQPPLPEREETGDEEDGSPIALHRGPPGSRGPLIPLLSLPPPPWG RGPIRRGLGPRSSPYGRGWGVNAEPFPGPHGGPTRGSFHKEQNPRLKSWSLIKNTCPP KDDPQVMEDKSDRPVCRHFAKKGHCYEDLCIFYHPGVNGPPL		

Further analysis of the NOV34a protein yielded the following properties shown in Table 34B.

Table 34B. Protein Sequence Properties NOV34a	
PSort analysis:	0.7000 probability located in nucleus; 0.2531 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV34a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 34C.

Table 34C. Geneseq Results for NOV34a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM93213	Human polypeptide, SEQ ID NO: 2614 - Homo sapiens, 167 aa. [EPI130094-A2, 05-SEP-2001]	66..232 1..167	166/167 (99%) 167/167 (99%)	e-105
AAU28194	Novel human secretory protein, Seq ID No 363 - Homo sapiens, 940 aa. [WO200166689-A2, 13-SEP-2001]	86..232 799..939	55/167 (32%) 63/167 (36%)	8e-13

AAU28382	Novel human secretory protein, Seq ID No 739 - Homo sapiens, 968 aa. [WO200166689-A2, 13-SEP-2001]	86..232 826..967	56/171 (32%) 61/171 (34%)	7e-12
AAM39141	Human polypeptide SEQ ID NO 2286 - Homo sapiens, 707 aa. [WO200153312-A1, 26-JUL-2001]	36..163 46..176	36/131 (27%) 48/131 (36%)	3e-05
AAM23916	Human EST encoded protein SEQ ID NO: 1441 - Homo sapiens, 1690 aa. [WO200154477-A2, 02-AUG-2001]	36..163 1222..1357	49/140 (35%) 55/140 (39%)	8e-05

In a BLAST search of public sequence databases, the NOV34a protein was found to have homology to the proteins shown in the BLASTP data in Table 34D.

Table 34D. Public BLASTP Results for NOV34a				
Protein Accession Number	Protein/Organism/Length	NOV34a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P79522	MHC class I region proline rich protein - Homo sapiens (Human), 253 aa.	1..232 1..253	232/253 (91%) 232/253 (91%)	e-145
Q96QB9	CAT56 protein - Homo sapiens (Human), 188 aa.	66..232 1..188	167/188 (88%) 167/188 (88%)	e-101
Q91Y15	Hypothetical 16.0 kDa protein - Mus musculus (Mouse), 143 aa.	101..232 12..143	113/132 (85%) 121/132 (91%)	7e-69
O55000	Hypothetical 92.8 kDa protein - Rattus norvegicus (Rat), 872 aa.	50..232 640..871	78/245 (31%) 94/245 (37%)	7e-16
AAH29765	Hypothetical protein - Mus musculus (Mouse), 623 aa (fragment).	50..232 375..622	78/261 (29%) 93/261 (34%)	4e-14

5

PFam analysis predicts that the NOV34a protein contains the domains shown in the Table 34E.

Table 34E. Domain Analysis of NOV34a			
Pfam Domain	NOV34a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-CCCH	200..226	10/27 (37%) 20/27 (74%)	0.0031

Example 35.

The NOV35 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 35A.

Table 35A. NOV35 Sequence Analysis			
	SEQ ID NO: 139	20566 bp	
NOV35a, CG119566-01 DNA Sequence	CCACTACTACTCTGAAAAATGGCAGATGACGAAGACTATGAGGAGGTGGTGGAGTACTACACA GAAGAAGTGGTTTACGAAGAGGTGCCGGGAGAGACAATAACAAAAATTTATGAGACTACGACA ACAAGGACATCTGACTATGAGCAATCAGAACTTCCAAACCAGCTCTGGCAGAGCCAGCACTG GCACAGCCAGCATCAGCAAAGCCGGTGGAGAGGAGGAAGGTCTCCGGAAGAAAGTGGATCCT TCAAAGTTCATGACCCCTACATTGCACACAGTCAGAAAATGCAGGATCTTTTATGCCCAAAT AAATACAAGGAGAAGTTTGAGAAAACAAAAGGACAGCCATACGCCAGCACAACAGATACTCCA GAACCTCGCAGAATCAAAAAAGTACAAGATCAACTCAGTGAGGTTAAGTATCGAATGGATGGT GATGTTGCTAAGACTATATGTCACGTAGATGAAAAGCAAAGGATATTGAACATGCAAAAGAAA GTGTGCGCAGCAAGTCAGTAAGGTTTATACAAGCAGAACTGGGAAGACACCAAGGATAAGTAC CTGCTTCCTCCTGATGCCCTGAACCTGTCCAGGCCGTAAAGAACACCGCCATGTCAGCAAG AAACGTGTACTGAAAGCTGGGAAGCAGACAAAAGTTGTTTTACCCCTATAATGATAGCCCG GAACTGAGGAGAGTTGCCAGGCCAGAAAGCTCTCAGTGATGTTGCCCTACAAAAAGGTCTC GCTGAACAGCAAGCTCAATTACAGCCTCTGGCCGATCCTCCAGATATAGAATTTGCCAAGAAA GTAACCAATCAAGTGAGCAAGCAAAAATACAAAGAAGACTATGAAAATAAATCAAAGGCAAA TGGAGTGAGACACCTTGCTTTGAAGTTGCAAATGCCAGAATGAATGCTGATAACATTAGCACA AGGAAATACCAGGAAGATTTTGAAGCATGAAAGACCAGATCTACTTCATGCAGACCGAAACA CCAGAGTATAAAATGAATAAAAAAGCTGGTGTGGCAGCTAGCAAGGTAAAAATACAAAGAAGAC TATGAAAAGAATAAAGGAAAAGCAGATTATAATGTGCTTCCTGCTTCAGAGAACCCACAGCTT AGGCAGCTGAAGGCAGCAGGAGATGCCCTAAGTGACAACTATACAAGGAAACTATGAAAAG ACAAAGCAAAGAGCATAAATTACTGCGAGACCCCAAATTCAGGCTCGATACGTGTTCTGCAG AACTTCAGTAGTGATAAAAAATATAAGATTCTACTTAAAGATATTTGGGACATTATGTA GGCAGCTTCGAGGATCCATACCAATTCACACTGCATGAAAGTCACAGCTCAAAACAGTGATAAA AACTACAAAGCAGAATACGAAGAAGACAGAGGCAAGGCTTCTTCCCTCAGACCATAACTCAA GAATATGAAGCAATTAAGAACTAGATCAGTGTAAGAGACCACACCTACAAAGTCCATCCAGAT AAGACAAAATTCACCAAGTTACAGACTCTCTGTTCTGCTACAAGCCCAAGTCAATTCCAA CAACTGAGTGACTTAAATTACAAAGCAAAACATGAAAGTGAAAGTTCAAGTGCCATATCCCC CCTGATACTCCTGCTTTTATCCAGCACAAGTCAATGCCTATAACTTGAGTGATAATCTTTAT AAGCAAGACTGGGAGAAGAGCAAGCCAAAAGTTTGACATTAAAGTGGATGCCATTCCCTGT CTGGCAGCCAAAGCCAACACCAAGAACACCAGCGATGTGATGTACAAGAAAGACTATGAAAAA AACAAAGGAAAAATGATTGGAGTCTCAGCATTAAATGACGATCCCAAGATGCTGACCTCTTG AAGGTGGCCAAAACAGAGTGATAGATTATACAAGGAAAACATAGAGAAGCAAAAGGCAAAAG AGTATGAATTACTGTGAGACCCCAAATATCAACTTGATACTCAGCTGAAGAACTTCAGTGAG GCTAGATATAAGACTTATATGTAAGGATGTTTTGGGACATTATGTAGGCAGCATGGAGGAC CCATATCACACACTGCATGAAAGTTGCAGCTCAAAACAGTGATAAAAGTTACAAGCAGAA TATGAAGAAGATAAAGGAAAATGCTATTTCCCTCAGACAATAACACAGAATATGACCAATC AAGAAGCTGGACCAGTGTAAGATCATACCTACAAAGTTCATCCAGATAAGACCAAATTCAG GCAGTCACTGATTCTCCTGTAATGTTGCAAGCCAGCTCAACACGAAACAGCTTAGTGATCTG AATTACAAAGCAAAACATGAAGGTGAGAGGTTCAAGTGCCATATACCAGCAGATGCTCCACAG TTTATCCAACACAGAGTCAATGCCTATAATCTGAGTGATAATGTTTATAAGCAAGACTGGGAG AAGAGCAAAGCCAAGAAGTTTGACATTAAAGTGGACGCCATTCCCCTGTTGGCAGCCAAAGCC AACACCAAGAACACCAGCGATGTGATGTACAAGAAAGACTATGAAAAGAGCAAAGGGAAAATG ATTGGAGCCCTCAGCATTAAATGACGATCCAAAGATGCTGCACTCCTTGAAAGCAGCCAAAAC CAGAGTGATCGGAATATCGAAAAGATTATGAAAAGTCAAAAATCTACACGGCACCTCTT GATATGCTCCAAGTCACTCAAGCTAAGAAATCTCAGGCAATTGCCAGCGACGTTGATTATAAG CACATCTTACACAGTTACAGCTACCCCCCTGATAGCATCAATGTGGACCTTGCCAAGAAGGCA TATGCGCTGCAGAGCGATGTTGAATACAAAGCTGACTACAATAGCTGGATGAAAGGTTGTGGC TGGGTGCTTTTGGGTCCTTAGAAATGGAAAAGGCAAGCGAGCTTCAGACATCCTCAATGAG AAAAAATATCGCAACATCCAGACACCTCAAGTTTACCTCGATTGAAGATGCTCCAATTACA GTACAGTCTAAATTAACAGGCCAGAGGAGTGATATCGCTTACAAAGCCAAAGGAGGAA ATTATTCACAATTACAACCTGCCACCAGACCTGCCCCAGTTTCATCCAGGCTAAAGTTAATGCC TACAATATCAGTGAGAATATGTACAAAGCAGACTGAAAGACTTGAGCAAGAGGGATATGAC CTGAGAACTGATGCGATTCCCATCAGAGCTGCCAAAGCTGCCAGGCAGGCGGAGTGACGTT CAGTACAAAAAGACTATGAAAAGGCTAAAGGGAAAAATGTTGGCTTCCAAGTCTTCAAGAT GACCCTAAACTGGTTCATTATATGAACGTGGCCAAGATACAATCAGATCGGGAGTATAAAAAA GACTATGAGAAGCAAAAGTCCAAATACAACACGCCCATGATATGTTCAATGTCGTGGCGGCT		

AAGAAAGCCAGGATGTGGTCAGCAATGTCAACTATAAGCATTCTCTCCATCATACACCTAC
TTGCTGACGCCATGGACCTGGAGCTGTCTAAGAACATGATGCAGATACAGAGTGATAACGTC
TACAAGGAAGACTACAACAACCTGGATGAAAGGCATTGGCTGGATTCTTATTGGCAGTCTCGAC
GTCGAAAAAGTTAAAAAGGCCGGTGATGCTCTGAATGAAAAGAGTACAGGCAACATCCAGAC
ACCTCAAAATTTACCAGCATTGTGGACTCCCCAGTTATGGTCCAGGCAAAACAGAACACGAA
CAAGTCAGTGATATCTTATACAAGGCTAAAGGAGAAGATGTGAAACATAAATACACCATGAGT
CCTGATCTTCTCAGTTTCTCAGGCCAAGTGCAATGCTTACAGTATAAGTGACGTCTGTTAT
AAACGGGATTGGCATGACTTAATACGCAAGGGCAACATGTGCTGGGCGATGCTATTCCCATC
ACTGCAGCCAAGGCATCGAGAAACATTGCCAGTGATTATAAATACAAGGAAGCTTATGAGAAG
TCAAAGGGAAGCATGTGGGTTTCAAGAGCCTCCAGGATGATCCCAAGCTGGTCCACTATATG
AATGTGGCAAAGCTGCAGTCTGATCGTGAATACAAGAAGAACTATGAGAACACCAAAACCAGC
TACCATACCCCTGGGACATGGTTACGATCAGAGCTCAAAGATGGCCAGGATGTGCTACC
AATGTCAACTACAAACAGCCATTGCATCATTACACATACCTACCTGACGCCATGAGTCTTGAG
CATACGAGGAATGTCAATCAAATTCAGAGTGATAATGTGTATAAAGACGAGTATAACAGCTTC
TTGAAGGGCATCGGATGGATCCCTATTGGTTCCCTGGAGGTGGAGAAGGTCAAGAAAGCAGGC
GATGCATTAAATGAGAGGAAGTATCGACAGCACCCAGATACCGTCAAGTTCACAAGTGTGCT
GATTCCATGGGCATGATGTTGGCTCAGCATAACACAAAGCAGCTAAGTGATTGAACTACAAG
GTAGAGGGAGAGAACTGAAGCACAAGTATATCTATTGACCTGAATTGCCTCAGTTTATTCAA
GCCAAAGTCAACGCCCTCAACATGAGTGATGCTCATTATAAAGCAGATTGGAAGAAAACCAT
CGCAAGGGCTATGATTGAGACCAGATGCCATCCCAATTGTTGCTGCAAAAAGTTCAAGGAAT
ATTGCTAGTGATTGCAAATATAAGGAGGCCTACGAGAAAGCCAAAGGCAAGCAAGTTGATT
CTCAGTCTTCAGGATGATCCTAACTGGTTCACTACATGAATGTGGCCAAAATCCAGTCTGAT
CGTGAGTACAAAAGGGCTATGAAGCCAGCAAGACCAAGTACCACACACCTCTGGATATGGTC
AGTGTGACAGCTGCAAAAGAAATCTCAGGAGGTGCCACCAACGCCAACTACAGACAGTCATAC
CACCCTACACTCTCTGCCCCGATGCTTGAATGTGGAGCACTCCAGGAATGCCATGCAGATT
CAGAGTGATAATCTGTACAAATCTGACTTCACCAATTGGATGAAAGGGATCGGCTGGGTGCCC
ATAGAGTCCCTGGAGGTGGAGAAGGCAAGAAAGCAGGAGAGATTCTTAGTGAGAAGAGTAT
CGCCAGCACCCCGAGAAGCTGAAGTTCATTACGCCATGGACACAATGGAACAGGCACCTTAAC
AAGAGTAACAACTGAACATGGACAAGAGGCTCTACACTGAAAAATGGAACAAGGACAAGACC
ACCATTCTATGTCATGCTGACACACCGGATATTTTACTCTCCAGAGTAAACCAAATCACCATG
AGTGATAAACTGTACAAAGCTGGCTGGGAAGAGGAAAAGAAGAAAGGATATGACCTGAGGCCT
GATGCCATTGCAATAAAGGCTGCAAGAGCCTCTAGAGACATTGCCAGTGATTACAAATACAAG
AAAGCCTATGAACAAGCCAAAGGGAACACATTGGCTTCCGGAGCCTGGAAGATGACCCCAAG
CTGGTGCACTTCATGCAAGTGGCCAAAGATGCAGTCAGACCGGGAATACAAGAGGGATATGAG
AAATCCAAGACCTCCTCCACACCCCGGTGGACATGCTCAGTGTGGTGGCAGCCAAGAAGTCT
CAGGAAGTGGCCACCAATGCCAACTACAGGAACGTGATCCATACCTACAACATGCTTCTGAT
GCCATGAGCTTTGAATTGGCCAAAAATATGATGCAGATTCAAAGTATAATCAGTACAAGGCT
GACTATGCTGACTTCATGAAGGGCATTGGATGGCTCCCTCTGGGCTCCCTGGAAGCAGAGAAA
AACAAGAAAGCCATGGAGATTATTAGTGAAAAGAGTACC GCCAGCACCCAGACACTTTGAAG
TATTCACACTCATGGAATCGATGAACATGGTTTTGGCCAGAATAATGCAAAAATTTATGAAC
GAACATCTCTACAAACAAGCATGGGAGGCTGACAAAACCAAAGTCCACATCATGCTGATATC
CCCCAGATTATTTGGCAAAGGCAAAATGCAATTAATATAAGTGATAAACTCTACAACTTTCC
TTGGAAGAGTCTAAAAAGAAAGGCTATGATCTCAGACCTGATGCAATTCCTATCAAAGCTGCC
AAGGCTTCCAGAGATATTGCAAGTGATTATAAATACAAGTACAATTATGAAAAAGGGAAGGGG
AAAATGGTTGGTTTCCGCACTCTCGAGGATGATCCAAATTAGTCCATTCCATGCAAGTGGCT
AAGATGCAATCTGATCGGAGTACAAGAAAACTATGAGAACACAAAGACCAGCTACCACACC
CCTGCCGACATGCTCAGTGTACGGCTGCAAGGATGCCAAGCCAACATCACCACACTAAC
TACAAGCACCTGATTACAAGTACATCCTCCTCCAGATGCAATGAACATTGAGCTGACCAGG
AATATGAATCGCATACAGAGTGATAATGAATATAAGCAAGATTACAATGAATGGTACAAAGGG
CTTGGCTGGAGTCCAGCAGGTTCTCTGGAAGTGGAGAAGGCCAAGAAAGCAACTGAATATGCC
AGTGATCAGAAATACCGCCAGCACCCGAGCAACTTCCAGTTTAAAGAGCTGACTGATTCCATG
GACATGGTGTCTGCAAGCAGAAATGCACATACCATGAACAAGCATTTATACACCATTGATTGG
AATAAAGATAAGACCAAGATTCTATGTGATGCCTGATACACCAGATATTTACAAGCCAAGCAG
AATCAAACACTGTATAGTCAGAACTCTATAAACTTGGATGGGAAGAAGCTTTGAAGAAAGGC
TATGATCTCCAGTTGATGCAATTTCTGTACAGCTAGCTAAAGCTTCAAGAGACATTGCTAGT
GATTATAAATACAAACAAGGCTACCGAAAGCAACTTGGCCACCATTGTTGGATTCCGGAGTCTG
CAAGATGACCCAAAACCTGTGTTGTCCATGAATGTAGCCAAAATGCAGAGTGAAAGAGAATAC
AAGAAGGACTTTGAGAAGTGAAAACCTAAGTTCTCCAGCCAGTGGACATGTTGGAGTGGTA
CTGGCCAAAGAGTGTAGGAGTTGGTTAGTGACGTGGACTACAAGAACTACCTGCATCAGTGG
ACATGTCTGCCTGATCAGAACGATGTTGTGCAAGCTAAGAAAGTTTATGAACTGCAAGTGAG
AATCTATATAAATCTGACCTTGAGTGGCTGAGAGGCATAGGATGGAGTCCCTTGGGTTCTTTA
GAGGCAGAAAAGAAACAGCGGCTTCGGAATCATCAGTGAGAAGAAATATCGTCAGCCTCCA
GACAGAAACAAGTTCAACAGCATTCCTGATGCCATGGATATAGTTCTGGCAAGACAATGGCC
AAAAATAGGAGTGATAGACTTTATAGAGAAGCTTGGGACAAAGACAAGACTCAGATCCACATC
ATGCCTGATACACCTGACATTGTTCTGGCTAAAGCAAACTTAATCAACACAAGTGATAAACTC
TACCGAATGGGTTATGAGGAGCTGAAGAGAAAAGGTTACGATCTTCTGTTGATGCCATACCA

ATCAAAGCAGCAAAAGCCTCCCGGAAATTGCCAGTGAATACAAGTACAAGGAAGGCTTTTCGC
AAGCAGCTCGGCCACCACATTGGTGCCCGAACATTGAAGATGACCCCAAGATGATGTGGTCC
ATGCATGTGGCCAAGATCCAGAGTGACAGGGAGTACAAGAAGGACTTTGAGAAGTGGAAAGACC
AAGTTCAGCAGCCCAGTGGACATGCTGGGGGTGGTGTGGCCTATAAGTGCCAGACCTTAGTC
AGCGACGTGGACTACAAGAACTACCTGCACCAAGTGGACATGCCTGCCCGACAGAGCGATGTC
ATCCATGCTCGGCAGGCCTATGACCTCCAGAGCGATAATTTGTACAAGTCAAGACCTTCAGTGG
CTAAAAGGCATTGGCTGGATGACTAGTGGTTCTCTCGAGGATGAGAAAAATAAACGAGCCACC
CAGATTTTGAAGTACCATGTTTACCGTCAGCACCCAGATCAATTTAAGTTTTCAGCCTTATG
GATTCATACCAATGGTTTGGCAAAAAACAATGCTATTACCATGAATCATCGCCTCTATACA
GAAGCTTGGGATAAAGATAAAACCACTGTCCACATTATGCCAGATACCCCTGAAGTTTATTA
GCTAAACAAAACAAGTAAATTACAGTGAGAAATTGTATAAGCTTGGCCTAGAAGAAGCCAAG
AGGAAAGGTTATGACATGCGGGTAGATGCCATTCCATCAAGGCAGCCAAGGCCTCCAGAGAT
ATTGCAAGTGAATTCAAGTACAAGAAGGCTATCGTAAGCAGCTCGGCCACCACATTGGTGCC
CGAGCTATACGTGATGACCCCAAGATGATGTGGTCCATGCACGTGGCCCAAGATCCAGAGTGAC
AGGGAGTACAAGAAGGACTTTGAGAAGTGGGAAGACCAAGTTCAGCAGCCCAGTGGACATGCTG
GGGGTGGTGTGCTGGCCAAGAGTGCCAGACCTTAGTCAGCGATGTGGACTACAAGAACTACCTG
CACCAGTGGACATGCCTGCCCGACAGAGCGACGTATCCATGCTCGGCAGGCCTATGACCTC
CAGAGCGATAATATGTACAAGTCTGATCTCCAGTGGATGAGAGGCATTGGCTGGGTGTCCATT
GGCTCTTTGGATGTGAAAAATGCAAAAGGCAACTGAAATTTTGAAGTGAATAAATCTATCGC
CAGCCTCCAGACAGATTCAAATTTACAGTGTGACTGACTCTCTGGAACAAGTGTGGCCAAG
AACAATGTCTCAACATGAATAAGCGTTTATACACAGAGGCCTGGGACAAAGACAAGACTCAA
ATTCACATAATGCCTGATACACCAGAGATTATGTTGGCAAGGCAGAACAAAATCAACTACAGT
GAGACTCTATACAACTTGGCAATGAAGAAGCAAAAAAGAAAGGCTACGACTTGGCAAGTGAC
GCCATCCCCATCGTGGCTGCCAAGGCCTCCAGGGACGTTATCAGTGATTACAAATACAAAGAT
GGTTACCGCAAGCAGCTCGGCCACCACATTGGAGCCCGGAACATTGAAGATGACCCCAAGATG
ATGTGGTCCATGCATGTGGCCAAGATCCAGAGTGACAGGGAGTATAAGAAGGACTTTGAGAAG
TGGAAGACCAAGTTCAGCAGCCCAGTGGACATGCTGGGAGTGGTGTAGCCAAAGAGTGCCAG
ACCTTAGTCAGCGATGTGGACTACAAGAACTACCTGCACGATGGACGTGCCCGACCCAG
AATGATGTATCCATGCTCGGCAGGCCTATGACCTCCAGAGCGATAACATTTACAAATCTGAT
CTCCAGTGGCTGAGAGGCATTGGCTGGGTCCCCATTGGGTCTATGGATGTGGTCAAGTGCAAG
AGAGCTGCTGAAATACTGAGTGATAACATCTACCGCCAGCCTCCGGACAAGCTGAAATTTACC
AGTGTGACTGACTCTCTAGAGCAGGTGCTGGCCAAAGAACAAATGCTCTCAATATGAACAAAGCG
TTATACACAGAAGCCTGGGACAAAGACAAGACCCAAAGTCCATATTATGCCTGATACACCTGAA
ATCATGTTGGCAAGACAAAATAAAATAAATTATAGTGAGAGCCTCTATCGTCAGGCCATGGAA
GAAGCCAAAGAAAGAGGCTATGACTTGAGAAGTGATGCCATTCCCATTGTGGCTGCCAAGGCC
TCTCGGGATATTGGCCAGTGATTACAAATACAAAGAAGCATATCGTAAGCAGTTGGGTACCAC
ATTGGCGCCCGAGCAGTACACGATGACCCCAAGATAATGTGGTCCCTCCACATTGCCAAAGTG
CAGAGTGACCGTGAGTACAAGAAAGATTTTGAGAAATACAAGACAAGGTACAGCAGCCAGTG
GACATGCTTGGTATCGTTTGGCCAAGAGTGTCAGACCTTGGTCAGCGATGTGGACTATAAA
CATCCTCTGCATGAATGCATCTGCTGCCCGACCAAGATGACATCATTCATGCACGGAAAGCC
TATGACCTCCAGAGTGACAAATTTGTATAAGTCAAGCCTTGAATGGATGAAGGCAATTGGCTGG
GTTCCGATTGATTCTTGAAGTTGTTAGGGCCAAGAGAGCTGGAGAATTACTTAGTGATACT
ATCTACCGTCAGCGTCCAGAAACGCTGAAATTTACCAGTATAACGGACACTCCGGAGCAGGTG
CTGGCAAAAAACAATGCTTTAAACATGAATAAGCGCTTATATACTGAAGCCTGGGACAATGAC
AAGAAAATCTATTGATGTCATGCTGATACACCAAGAAATCATGTTAGCCAAATCAACCGAATA
AACTACAGTGATAAACTCTATAAATCTGCTTTGGAAGAGTCCAAGAAGGAAGGCTATGACTTG
CGTCTGGATGCCATTCCAATCCAGCAGCCAAGGCTTCAAGAGATATTGCTAGTGATTACAAG
TACAAGGAAGGCTACCGCAACAGCTTGGCCACCATATTGGGGCCCGGAACATTAAAGGATGAC
CCGAAGATGATGTGGTCCATCCATGTGGCCAAGATCCAGAGTGACAGGGAGTACAAGAAGGAG
TTTGAGAAGTGAAGACCAAGTTCAGCAGCCAGTGGACATGTGGGGGTGGTGTGGCCAAAG
AAGTGTGAGATCCTTGTAAAGCAGATAGACTACAAGCATCCCCTGCATGAATGGACCTGCCTG
CCTGATCAGAAAGACGTCATTAGGCTCGGAAGGCCTATGACCTGCAGAGTGATGCTATTTAC
AAATCTGATCTTGAGTGGCTGAGAGGCATAGGATGGGTCCCATTGGCTCTGTAGAGGTCGAG
AAAGTGAAGAGAGCTGGAGAAATCCTGAGTGACAGGAAGTATCGCCAGCCTGCAGACCAAGCT
AAATTCACATGCATTACCGACACTCCGGAAATTGCTCTAGCAAAGAATAATGCCCTGACAATG
AGCAAGCATTTATACACAGAAGCTTGGGATGTGACAAAACCTCCATCCAGTGATGCCAGAC
ACCCAGATATCTGCTGGCCAAAGAGTAATTTGCCAATATCAGCCAAAAACTTTACACCAAG
GGATGGGATGAATCAAGATGAAGACTATGATCTGAGAGCAGATGCTATTTCCATCAAAAGT
GCCAAGGCCTCAGGGACATCGCCAGTGACTACAAATACAAGGAAGCCTATGAGAAACAGAAA
GGCCACCACATTGGAGCCAGAGCATTGAAGATGATCCCAAGATTATGTGTGCCATACATGCA
GAAAAAATTCAAAGTGAAGGGAGTACAAGAAGGAATTCAAAAGTGGAAACCAAGTTCTCT
AGCCCAAGTGACATGTTAAGCATCTTGTCTGGCCAAGAAATGTGAGACTTTGGTCACTGACATT
TATTATCGCAATTACCTGCATGAATGGACATGATGCCGATCAAAACGACATTATCCAAAGCA
AAAAAGGCCTATGACCTGCAGAGTGATGCCCTCTACAAGGCTGACTTGGAGTGGTGTGGTGGC
ATTGGCTGGATGCCCCAAGGGTCTCTGAAAGTGTGAGAGTCAAAACGCCCAGAAATATCTTT
TGTGACAGTGTCTATCGGACGCCTGTGGTGAACCTTAAGTACACAAGCATTGTTGACACACCT

GAAGTGGTCTTGCTAAATCAAATGCTGAAAATATTAGTATTCCAAAGTACAGAGAGGTTTGG
GACAAGGATAAACTTCAATACACATAATGCCAGATACTCCAGAAATTAATCTCGCTAGAGCA
AATGCTCTTAATGTGAGCAATAAACTTTACCGTGAGGGCTGGGATGAAATGAAGGCGGGCTGT
GATGTCCGGCTGGATGCCATCCCCATCCAGGCTGCCAAGGCTCCAGGGAGATTGCCAGTGAC
TATAAATATAAGCTTGACCATGAGAAGCAGAAGGGACACTACGTGGGCACCTCACAGCCAGG
GATGACAACAAGATCCGCTGGGCCCTCATAGCTGACAAGCTCCAGAATGAACGAGAGTACCGG
CTGGACTGGGCCAAATGGAAGGCCAAGATCCAGAGCCCTGTGGACATGCTTTCCATCCTGCAC
TCTAAAAATTTCCAGGCTCTGGTCACTGACATGGATTACCGCAATTACCTGCACCAAGTGGACC
TGCATGCCCGACCGAAGCATGTGATTAGGCCAAGAAGGCTACGAACTGCAGAGCGATAAT
GTTTACAAGGCTGACTTGGAAATGGTTGCGTGGAAATGGGTGGATGCCAAATGACTCCGTGTCC
GTCAATCATGCCAAACATGCCCGGACATCTTCAGTGAGAAAAATATCGCACAAAAATAGAA
ACTCTCAACTTTACGCTGTGGATGACAGAGTTGATTATGTGACAGCGAAACAAAGTGGCGAG
ATCCTCGATGATATTAATACCGGAAAGACTGGAATGCCACCAATCAAAGTACACCTCACA
GAAACCCCTGCTGCACACTGCCAGGAGCTGCTAGGATACCTGGACCATCTCTTACAAAG
GAAGGCTGGGAGAGACAAAAAGCCACAGGTTACATTTGCCCTCCAGATGCTGTGCCATTGT
CATGCCCCATCACTGCAATGACGTTTCAAGTGAGCTGAAATACAAAGCTGAACATGTGAAGCAA
AAAGGTCAATTATGTTGGTGTCCGACGATGAGAGATGATCCTAAGCTGGTTTGGTTGAGCAT
GCAGGCCAGATTGAGAATGAGAGACTATACAAAGAGGACTATCACAAACAAAGGCCAAATC
AATATACCTGCTGATATGGTGTCACTCTTGCCCGCAAGCAGGGGCAGACCTGTGTCACTGAT
ATTGATTATCGTAATTACTTGCACCAATGGATGTGTCTCTGACCGAAGCATGTTATTTCAG
GCAAGAAAGGCTATGACCTACAGAGTGATAATGTCTACAGAGCTGACCTGGAGTGGTCCGA
GGCATTGGCTGGATCCCACTGGATTCTGTGGACCATGTAAGGGTTACTAAGAACAGGAAATG
ATGAGTCAGATCAAATATAAGAAAAATGCCCTTGAAAACTATCCTAACTTTTACAAGTGTGGT
GATCCTCCAGAGATTGTTTTAGCCAAGATTAATTCTGTCAATCAAAGTGATGAAAAATATAAA
GAAACATTTAATAAAGCAAAGGGCAAATATACGTTTTCCAGAGATACACCATATCTCCAC
TCCAAAGACATGGGAAAACTCTACAGTACTATCTGTATAAAGGGGCGTGGGAGGGCACCAAG
GCCTATGGCTACACCTGGATGAGCGCTACATTCCTATTGTGGAGCCAAGCATCTGTATCTG
GTGAACAGTGAGCTTAATACAAAGAGACATATGAGAAGCAGAAAGGTCACTACCTGGCTGGA
AAAGTGATCGGTGAATTCCTGGTGTGGTTCACTGTCTGGATTTCAAAAGATGAGGAGTGG
TTGAACTACAGAAACATTATGAGGATACCAAAGCAAATGTTTATATCCCAATGACATGATG
AATCAGTGTGCTGCTAAAGGTGCCAGTACATCCTCAGTGACCTGGAGTATCGACACTATTT
CACCAGTGGACGTCTCTCTGGAAGAACCAATGTTATACGCGTCCGAAACGCCAGGAGATC
TTGAGTGATAATGTGTATAAAGATGACCTGAATTGGTTGAAAGGCATTGGTTGTACGTTTGG
GATACACCCCAATCCTCCATGCCAAGAAATCATACGACCTTCAGAGTCAGTCAATATACA
GCAGCAGGTAAGAAAACTACAAAACTATAATCTGGTCACAGACACGCCCTCTATGTGACT
GCTGTTTCAAGAGTGGCATTAAATGCCAGTGAGGTAAAAATATAAAGAAATATCATCAGATTAA
GACAAATACACAACAGTTCTAGAAACAGTGGATTATGACAGAACCAGAAACCTGAAGAATCTT
TACAGCAGTAACCTGTACAAGGAGGCTGGGATAGAGTGAAAGCCACAGCTACATCCTGCCT
TCCAGCACCTTGTCCCTGACACACGCCAAGAACCAGAAGCATCTGCCAGCCATATCAAATAT
CGGGAAGAATATGAAAGTTCAAAGCTCTTTATACGTTACCAAGAAGTGTGACGATGATCCG
AACACAGCAGGCTGCCTCCGAGTTGGCAAGCTTAACATCGATCGCCTGTACAGATCAGTTTAT
GAAAGAACAAAGATGAAATCCACATCGTGCCCGACATGGTAGAGATGGTTACTGCCAAGGAT
TCCAGAAAGAAAGTCAGTGAGATTGATTACCGCTGCGCTCCACGAATGGATTGGCCACCCC
GACTTGCAAGTCAATGATCAGTCAGGAAAGTCACAGATCAGATCAGCGATATGTATACAAG
GATGACCTCAACTGGCTGAAAGGCATTGGTTGCTACGTCTGGGACACTCTGATAATCTCCAT
GCCAAGCATGCTTATGATCTACGTGATGATATCAAGTATAAAGCTCACATGTTGAAAAACAAG
AATGACTACAAGCTTGTACAGATACACAGTCTACGTGAGGCTGTCAAAGTGGGAAACAG
CTAAGTGACGCTGTCTACCACTATGACTATGTGCACAGTGTGAGAGGCAAAGTGGCTCCAAT
ACCAAAACCGTGGATCTGGACCGGGCCCTTCATGCATACAAGCTCCAGAGTTCGAATCTATAC
AAAACAGCCTGCGCACCTGCCCCTGATATAGACTTCCAGGTGACACTCCTCACTTCAAA
CACATCAAGGACACCCGTTACATGAGCAGTTATTTCAAGTACAAAGAAGCCTATGAACACACC
AAGGCATATGGGTATACACTTGGCCCCAAAGATGTTCCATTTGTCCACGTCGGGAGAGTCAAC
AATGTTACCAGCGAGAGACTGTATCGGGAATTGTACCACAACTGAAAGACAAGATCCATACA
ACTCCGATCCCCCTGAGATCCGCCAAGTCAAGAAGACACAAGAGGCTGTGATGAGTTGATC
TACAAATCAGACTTCTTCAAGATGCAGGGCCACATGATCTCTCTGCCATACACCCCCAAGTG
ATCCATTGCCGCTATGTGGGAGACATCACCAGTGATATTAATACAAAGAGGACTTGCAGGTC
CTGAAGGGATTGGCTGCTTCTGTATGACACTCCTGACATGGTCCGCTCCCGGACCTGCGG
AAGCTCTGGTCTAATTACCTATACACTGATAAGGCAAGGGAGATGCGAGACAAATACAAAGTG
GTGCTTGACACTCCAGAATACAGAAAGTGCAGAAGTGAAGACACATCTGAGTGAGTGGTC
TACAGAGCTGCAGGCAAGAAGCAGAAGTCAATCTTACTTCAGTTCCTGATACTCCTGATCTT
TTAAGAGCCAGCGAGGGCAGAAGCTTCAGAGTCAGTATCTGTATGTTGAACTTGCCACCAAA
GAGAGACCCCATCATCAGCTGGAACCCAGACCACAGCCTTGAAGCATGCTAAAGACGTTGAAG
GACATGGTCAGTGAGAAAAAGTACAAGATTCAATATGAAAAGATGAAAGACAAGTACATCCG
GTTCCAGATACGCCAATCCTCATCAGAGCCAAGAGGGCTTACTGGAATGCCAGTGATCTACGC
TACAAAGAAACATTTCAAAGACCAAGGGAAATACCACAGGTGAAAGATGCCCTAGACATT
GTCTATCATCGAAAGTCACAGATGACATCAGTAAAAATAAATACAAGGAGAACTACATGAGC

CAGTTGGGTATCTGGAGGTCCATTCTCGATCGTCCAGAGCATTTCACCACCGAGCAGTCACT
GACACAGTCAGTGATGTAAATATAAGAAGACTTGACTTGGCTTAAAGGCATTGGTTGCTAT
GCCTATGATACCCCTGATTTCACTCTGGCTGAAAAGAACAGACTCTCTACAGCAAGTATAAG
TATAAAGAAGTATTTGAAAGGACAAAGTCAGATTTCAAGTATGTTGCCGACTCTCCGATCAAT
AGGCATTTCAAGTATGCAACTCAATTGATGAATGAGAAAAATACAGAGCTGATTATGAGCAG
CGGAAAGATAAATACCACCTGGTAGTCGATGAGCCTAGACATCTGCTGGCTAAGACCCGACG
GACCAGATCAGTCAGATCAAATACAGGAAAACTATGAAAAATCAAAGGACAAATTTACCTCA
ATTGTGGATACTCCAGAACCTGCGTACTACAAAAGTCAACAAACAAATCAGCGATATCCTT
TATAAATTGGAATACAACAAGGCCAAACCCAGAGGCTACACCACAATCCAGACACGCCCATG
TTGCTGCTGTCCGCAAGGTTAAAGATGAAGTCAGTGATCTGAAATACAAAGAAGTATACCAA
AGAAATAAATCCAATGCACCATGAGCCAGATGCTGTTTATATCAAAGCAGCAAGGACGCC
TACAAAGTCAACACCAATCTGGACTATAAGAAACAGTACGAAGCCAAACAGGCCACTGGAAG
TGGACTCCTGACCGACCGGACTTCTCCAGGCTGCCAAGTCATCCCTGCAGCAAGCGATTTT
GAATATAAGCTGGACCGGAGTTCTCAAGGTTGCAAGCTTCTGTCTGATGACAAAAAC
ACGGTGTCTGCCCTCAGGAATACTTAAATAGAAAGTGATCTGAAATACAAAGAGAAACATGTC
AAGGAAAGAGGAACCTGTGATGCCGTACCTGACACGCTCAGATCCTGCTGGCGAAGACTGTC
AGCAACCTGGTGTCTGAGAACAGTACAAGGACCATGTCAAGAAGCACTTGGCACAGGGCTCA
TACACAACACTACCAGAGACCGGACACTGTTTACGTCAAGGAAGTGACCAATGTCAGT
GATACAAATTACAAAAGAAGTTTGTCAAGGAGAAAGGAAAAATCCAATCTCTCATGCTG
GAGCCACCAGAGGTGAAACATGCTATGGAAGTGGCCAAGAAGCAAAGTGATGTCGCTTACAGA
AAAGATGCCAAAGAGAAGCTGCATTACACCACAGTGGCTGATCGACAGACATCAAGAAGGCC
ACACAGGCAGCCAAACAGGCCAGTGAGGTGGAGTACAGAGCCAAGCACCGCAAGGAAGGCAGC
CATGGCTTAAGCATGCTCGGTGCCCGACATAGAAATGGCCAAGAAGGCAGCAAGCTGAGC
AGCCAGGTAAATACCGAGAAAATTTGATAAAGAAAAGGGCAAGACACCAAAATACAAATCCA
AAAGACAGCCAGCTCTACAAAGTATGAAAGATGCTAATAATCTTGAAGTGAGGTTAAATAC
AAGGCTGACCTGAAGAACTTACAAACCCGTGACTGACATGAAGGAGTCTCTGATCATGAAT
CATGTCTGAATACAAGCCAATTGCCAGTCTTACCAGTACAAGAAGATGAGTGAAGT
AAAGGCCACTACCACACCATAACCGATAATCTGGAGCAGCTTACCTAAAAGAGGCCACAGAA
TTACAGAGTATAGTGAATACAAAGAAAAGTATGAAAGGAACGAGGAAAACCCATGCTGGAC
TTTGAAACACCAACGTACATCACTGCCAAAGAGTCTCAGCAGATGCAGAGTGGGAAAGAAATAT
AGGAAAGATTATGAAGAGTCCATTAAAGGCAGAAACCTGACTGGCCTGGAGGTCACGCCAGCT
TTGTTACATGTCAAATATGCAACTAAAATAGCAAGCGAGAAAGAGTACAGGAAAGATCTGAG
GAAAGCATCCGTGGGAAGGCCCTCACTGAAATGGAAGATACACCTGACATGCTAAGAGCAAAG
AATGCCACTCAAATCCTCAATGAGAAAGAAATATAAGCGAGACCTGGAAGTGAAGTCAAAGGA
AGAGGCCCTGAATGCCATGGCCAAATGAACTCCGGATTTTATGAGGGCCAGGAATGCTACTGAT
ATTGCCAGTCAGATTAAGTATAAGCAATCAGCAGAAATGGAGAAAGCCAATTTCACTTCTGTG
GTTGATACTCCAGAGATCATTCATGCCCAACAAGTCAAGAATCTTTCAAGCCAGAAAAAGTAC
AAGGAAGATGCTGAGAAGTCCATGTCGATTATGAGACTGTTTGGACACCCAGAGATACAG
AGAGTCCGGGAGAACCAAGAACTTACGCTTCTCCAATACCAGTGTGACCTTAAACACAGT
AAAGGAAAAATTACAGTTGTTTCAAGACACGCCAGAAATCTGCGTGTAAAAGAAAATCAGAAG
AATTTAGCTCGGTTTTATATAAGAGGATGTCCTACCAGGAACCGCTATCGGAAAGACACCT
GAGATGATGAGAGTGAACAAACACAGGACCACATTAGCTCGGTGAAGTATAAGGAAGCAATA
GGACAAGGAATCCAATCCTGACCTGCCTGAAGTGAACGCTGTGAAGGAGACGCAGAAGCAC
ATTAGCTCGGTTATGTACAAAGAAAACCTGGGAACAGGAATTCATCCCATCACTCTCTGAG
ATTGAGAGAGTCAACACAATCAAGAAAACCTTAGCTCGGTGTTATACAAAGAAAACATGGGC
AAGGGAACCCCTTTACCTGTCACTCCCGAGATGGAAAGAGTCAACACAATCAAGAAAATATT
AGCTCGGTGTTATACAAAGAAAACATGGGCAAGGGAACCCCTCTACCTGTCACTCTCTGAGATG
GAGAGAGTCAACACAATCAAGAAAATATTAGCTCGGTGTTATACAAAGAAAACATGGGCAAG
GGAACCTCTTTACCTGTCACTCCCGAGATGGAGCGAGTCAACACAATCAAGAAAATATTAGC
TCGGTTTTGTACAAAGAAAATGTGGGGAAGCCACCGCAACCACTGTGACTCTGAGATGCAG
AGAGTCAAACGCAATCAAGAAAACCTTTAGCTCGGTGTTATACAAAGAGAACTGGGGAAGCA
ACCCCAACACCCCTTTACTCCCGAGATGGAGCGAGTCAAACGCAATCAAGAAAATATTAGCTCG
GTTTTGTACAAAGAGAACATGAGAAAAGCAACTCCGACACCTGTTACTCCAGAGATGGAGAGA
GCTAAGCGCAACCAAGAAAACATTAGCTCGGTTCTTTATCTGATAGTTTCCGGAACAAATA
CAAGGCAAAGCTGCCTATGTATTGGATACCCCGAGATGAGACGGGTGAGGGAGACCCAAACGG
CACATCTCAACGGTGAATATCATGAAGACTTTGAGAAACACAAGGGTGTCTTACACCAGTG
GTGACAGATCCTATCACTGAACGAGTAAAGAAGAACATGCAGGACTTCAGTGACATTAACCTAC
CGAGGTATTAGAGGAAAGTGGTAGAAATGGAACAAAACGGAATGACCAAGATCAGGAGACT
ATTACAGGTTTACGTGTCTGGCGTACTAATCCTGGTTCGGTTTTTGAATGATCTCCAGCAGAA
GACAACATCCAGTCCCGAAGCTTACACATGATTAATGTCCAAGCTCAGCGCCGAGCCGGGAG
CAGTCACGATCTGCCAGTGCCTAAGCGTCAGTGGGGGTGAGGAGAAGTCTGAGCATTACAGAA
GCACCAGACCACACCTTTGACTTACAGCGACGGGGGTGTCTTTGCAGTCTCAACAGCTTAC
AAACATGCAAAAACCAAGAGCTCCCAACAACAGATCATCTTCAAGTGTACCAACAGACA
ACGGTATCTTCCATCCCATCTCATCCATCTACTGCTGGAAAAATCTTCCGTGCCATGTATGAC
TATATGGCTGCTGATGAGATGAGGTGCTCTTCAAGGATGGAGATGCCATCATAAATGTTCAA
GCAATTGATGAAGGCTGGATGTATGGCACTGTGCAGAGGACTGGCAGGACCGGAATGCTCCCA

	GCCAACTACGTTGAAGCTATTAGGCATTTCAAAGCATCACACTTGTCTGCAGGACTTACAGA TCCTGCAGTCAATGTTTCGGTTTAGACTCTCCACTGTACCTAAGTTCTCAAGCTGCCTATGG TTTTTCTGTGTCAATGTGATTATGGTAGTACCATCCTTTCTCCTTTGGGTTTTAAATAAGT TGCAGAACAGACACTTTAAAGCTTCTGCAATATTATTCTGTGCCTAGAGTCTTTCTCCATT ATAAACATGTTTTAACATTATTTCTTTTCTAAACAGGGATTGTAATATGCCAACACATTA AAGGAAAAATAGCAGAGATGTTACCTTTTCTTGTCTGATTGCTAATGCTTATTATTTCTAAT TCAGTTCTGAAGTTATAAACTTATAATCAATACAAACAGCACTAATAAAACCTCTAATTCT GCAAAAAAAAAAAAAAAAAAAAAAGTCG		
	ORF Start: ATG at 19		ORF Stop: TAG at 20119
	SEQ ID NO: 140	6700 aa	MW at kD
NOV35a, CGI19566-01 Protein Sequence	MADDEDYEEVVEYYTEEVVVEVPGETITKIYETTTTRTSDEYQSETSKPALAQPALAQPASA KPVERRKVIKKVDPKSFMTPIYIAHSQKMQDLFSPNKYKEFEKTKGQPYASTTDTPELRRRIK KVQDQLSEVKYRMDGDVAKTICHVDEKAKDIEHAKKVSQVSKVLYKQNWEDTKDKYLLPPDA PELVQAVKNTAMFSKKLYTEDWEADKSLFYFYNDSPELRRVAQAQKALSDVAYKKGLAEQQAQ FTPLADPPDIEFAKKVTNQVSKQKYKEDYENKIKGKWSSETPCFEVANARMNADNISTRKYQED FENMKDQIYFMQTETPEYKMNKKAGVAASKVKYKEDYEKNKGKADYNVLPASENPQLRQLKAA GDALSDKLYKENYEKTKAKSINYCETPKFKLDTVLQNFSSDKKYKDSYKLDILGHYVGSFEDP YHSHCMKVTAQNSDKNYKAEYEEDRGKGFPPQITITQEYEAIKKLDQCKDHTYKVVHPDKTKFTQ VTDSPVLLQAQVNSKQSLDNLNYKAKHESEKFKCHIPPDTPAFIQHKVNAYNLSDNLKYQDWEK SKAKKFDIKVDAIPLLAAKANTKNTSDVMYKKDYKNGKMGVLSINDDPKMLHSLKVAKNQ SDRLYKENYEKTKAKSMNYCETPKYQLDTQLKNFSEARYKDYLVKDVGLGHYVGSMEDEPYHTHC MKVAAQNSDKSYKAEYEEDKGCYFPQITITQEYDAIKKLDQCKDHTYKVVHPDKTKFTAVTDSP VLLQAQNLTKQLSDNLNYKAKHEGERFKCHI PADAPQFIQHRVNAYNLSDNVYKQDWEKSKAKK FDIKVDIPLLAAKANTKNTSDVMYKKDYKSKGMIGALSINDDPKMLHSLKTAQNQSDREY RKDYKESKTIYTAFLDMLQVTQAKKSQAIASDVVDYKHILHSYSYPPDSINVDLAKKAYALQSD VEYKADYNSWMKCGGWVPPGSLMEKAKRASDILNEKKYRQHPDTLTKFTSIEDAPITVQSKIN QAQRSDIAYKAKGEEIHNYNLPDLPPQFIQAKVNAYNISENMYKADLKLKSKGYDLRTDAI PIRAAKAAQAAASDVQYKKDYKAKGMVGFQSLQDDPKLVHYMNVAKIQSDREYKKDYKTK SKYNTPHDMFNVVAAKKAQDVSNVNYKSLHHTYLPDAMDLELSKNMMQIQSDNVYKEDYN NWMKGIGWIPIGSLDVEKVKKAGDALNEKKYRQHPDTLTKFTSIVDSPVMVQAKQNTKQVSDIL YKAKGEDVKHKYTMSDPLPQFLQAKCNAYSISDVCYKRDWHDLIRKGNVNLGDAIPITAAKAS RNIASDYKYKAEYKSKGKHVGFRLQDDPKLVHYMNVAKIQSDREYKKGYEASKTKYHTPLDMVSVTAAK KAGEILSEKKYRQHPDKLFTYAMDTEQALNKSNNKLNMDKRLYTEKNWCKDTTIHVMP DTPDILLSRVNQITMSDKLYKAGWEEKKKGYDLRPAIPIKAARASRDIASDYKYKAYEQ KGKHIGFRSLEDDPKLVHFMQVAKMQSDREYKKGYEKSSTSFHTPVDMLSVVAAKKSQEVATN ANYRNVITHYNNLPDAMS FELAKNMMQIQSDNQYKADYADFMMKIGWPLGLSLEAKNKKAME IISEKKYRQHPDTLKYSTLMSNMVLAQNNAKIMNEHLYKQAWEDKTKVHIMPDIPIILA KANAINISDKLYKLSLEESKKKGYDLRPAIPIKAASRDIASDYKYKYNYEKKGKGMVGF SLEDDPKLVHSMQVAKMQSDREYKKNYENTKTSYHTPADMLSVTAAKDAQANITNTNYKHLIH KYILLPDAMNIELTRNMNRIQSDNEYKQDYNEWYKGLGWSFAGSLEVEKAKKATEYASDQKYR QHPSNFQFKKLTDSMDMVLAKQNAHTMKNHLYTIDWNKDKTKIHVMPDTPDILQAKQNTLYS QKLYKLGWEEALKGYDLPVDAISVQLAKASRDIASDYKYKQYRQQLGHHVGFRLQDDPKL VLSMNVAKMQSERYKKDFEKWTKFSSPVDMLGVVLAKKQELVSDVDYKYNLHQWTCPLDQ NDVVQAKKVYELQSENLYKSDLEWLRGIGWSPLGSLEAEKNKRASEIISEKKYRQPPDRNKFT SIPDAMDIVLAKTNAKNRSRLYREAWDKDKTQIHIMPDPDILVAKANLINTSDKLYRMGYE ELKRKGYDLPVDAIPIKAASREIASEYKYKEGFRKQLGHHIGARNIEDDPKMMWSMHVAKI QSDREYKKDFEKWTKFSSPVDMLGVVLAYKQCTLVSDVDYKYNLHQWTCPLDQSDVIHARQA YDLQSDNLKSDQLWLKIGWMTSGSLEDEKNKRATQILSDHVYRQHPDQFKFSSLMDSIPMV LAKNNAITMNRHLYTEAWDKDKTIVHIMPDPTEVLLAKQNKVNYSEKLYKLGLLEEAKRKGDM RVDAIPIKAASRDIASEFYKGYRQQLGHHIGARAI RDPKMMWSMHVAKIQSDREYKKDFEKWTKF FEKWKTKFSSPVDMLGVVLAKKQCTLVSDVDYKYNLHQWTCPLDQSDVIHARQAYDLQSDNMY KSDLQWMRGIGWVVISGLDVECKRATEILSDKIYRQPPDRFKFTSVTDSLEQVLAKNNALNM NKRLYTEAWDKDKTQIHIMPDPTEIMLARQNKINYSLEYKLANEAKKGYDLRSDAIPIVA AKASRDVISDYKYKDYRKQLGHHIGARNIEDDPKMMWSMHVAKIQSDREYKKDFEKWTKF SPVDMLGVVLAKKQCTLVSDVDYKYNLHEWTCPLDQNDVIHARQAYDLQSDNIYKSDLQWLRG IGWVPIGSMVDVVKCRAAEILSDNIYRQPPDKLFTSVTDSLEQVLAKNNALNMNRLYTEAW DKDKTQVHIMPDPTEIMLARQNKINYSLEYRQAMEEAKKEGYDLRSDAIPIVAASRDIA DYKYKAYRQQLGHHIGARAVHDDPKIMWSLHIAKVQSDREYKKDFEKYKTRYSSPVDMLGIV LAKKQCTLVSDVDYKHLHECICLPDQNDI IHARKAYDLQSDNLKSDLEWMMKIGWVPIDSL EVVRAKRAGELSDTIYRQRPETLTKFTSITDTPQVLAKNNALNMNKRLYTEAWNDKDKTIHV		

	MPDTP EIMLAKLNRINYSKLYKLAL EESKKEGYDLRLDAIPIQAAKASRDIA SDYKYKEGYR KQLGHHIGARNIKDDPKMMWSIHVAKIQSDREYKKEFEKWKTKFSSPVDMLGVLAKKQCILV SDIDYKHPLEHWETCLPDQNDVIQARKAYDLQSDAIYKSDLEWLRGIGWVPISGEVEKVKRAG EILSDRKYRQPADQLKFTCITDTP EIVLAKNNAL TMSKHYL TEAWDADKTSIHVMPDTPDILL AKSNSANISQKLYTKGWDESKMKDYDLRADAI SIKSAKASRDIA SDYKYKEAYEKQKGHHIGA QSIEDDPKIMCAIHAEKIQSEREYKKEFQKWKTKFSSPVDMLSILLAKKQCTLVTDIYYRNYL HEWTCMPDQNDIIQAKKAYDLQSDALYKADLEWLRGIGWMPQGSPEVLRVKNAQNI FCDSDVYR TPVVNLKYTSIVDTP EIVLAKSNAENISIPKYREVWDKDKTSIHIMPDTPEINLARANALNVS NKLYREGWDEMKGACDVRDLDAIPIQAAKASREI ASDYKYKLDHEKQKGHYVGTLTARDDNKIR WALIADKLQNEREYRLDWAQWKAKIQSPVDMLSILHSKNSQALVSDMDYRNYLHQWTCMPDQN DVIQAKKAYELQSDNVYKADLEWLRGIGWMPNDSVSVNHAKHAADIFSEKKYRTKIETLNFTP VDDRVDYVTAQSGEILDDIKYRKDWNATKSKYTLTETPLHLHTAQEAARILDQYLYKEGWERO KATGYILPPDAVPFVHAHCNDVQSELKYAEHVQKQGHYVGVPTMRDDPKLVWFEHAGQIQN ERLYKEDYHKTAKINI PADMVSVLAAGQGTLVSDIDYRNYLHQWMPCHPDQNDVIQARKAYD LQSDNVYRADLEWLRGIGWIPLDSVDHVRVTKNQEMMSQIKYKNALENYPNFTSVVDPPEIV LAKINSVQSDVKYKETFNKAKGKYTFSPDTPHISHKDMGKLYSTILYKGAWEGETKAYGYTL DERYIPIVGAHADLVNSELKYKETYEKQKGHYLAGKVIGEPGVVHCLDFQKMRSA LNYRKH YEDTKANVHIPNDMMNHVLA KRCCYILSDLEYRHYFHQWTSLEEPNVIRVRNAQEILSDNVY KDDLNLWLGIGCYVWDTPQILHAKKSYDLQSQLQYTAAGKENLQNYNLVTDTPLYVTAVQSGI NASEVKYKENYHQIKDKYTTVLETVDYDRTRNLKNLYSSNLYKEAWDRVKATSYILPSTLSL THAKNQKHLASHIKYREEYEKFALYTLPRSVDDDPNTARCLRVGKLNIDRLYRSVYEKNMKM IHI VPMVMEMVTAQDSQKKVSEIDYRLRLHEWICHDPDQNDHVRKVTDQISDVIYKDDLNLWL KGIGCYVWDTP EILHAKHAYDLRDDIKYKAHMLKTRNDYKLVTDTPVYVQAVKSGKQLSDAVY HYDYVHSVRGKVAPT TKTVDLRALHAYKLQSSNLYKTSRLTPTGYRLPGDTPHFHKIKDTR YMSSYFYKYAEAYEHTKAYGYTLGPKDVPFVHVRVNVNTSERLYREL YHKLKDKIHTTDPDPE IRQVKKTQEA VSELIYKSDFFKMQGHMISLPYTPQVIHCYVGDITS DIKYKEDLQVLKGFGC FLYDTPDMVRSRHLRLKWSNLYLTDKAREMRDKYKVVLDTPEYRKVQELKTHLSELVYRAAGK KQKSIFTSVPDTPDLLRAKRGQKLQSQYLYVELATKERPHHHAGNQTTALKHAKDVKDMVSEK KYKIQYEMKMDKYTPVPDTPILIRAKRAYWNASDLRYKETFQKTGKYHTVKDALDIVYHRKV TDDISKIKYKENYMSQLGIWRSIPDRPEHFHRAVTDTVSDVKYKEDTLWLKGIGCYAYDTPD FTLAENKNTLYSKYKYEVEFERTKSDFKYVADSPINRHFKYATQLMNEKKYRADYEQRKDKYH LVVDEPRHLLAKTRSDQISQIKYRKNYEKS KDKFTSIVDTP EHLRTTKV NKQISDILYKLEYN KAKPRGYTTIHDT PMLLHVRKVKDEVSDLYKEVYQRNKS NCTIEPDAVHIKAAKDAYKVNTN LDYKKQYEANKAHWKWTPDRPDFLQAAKSSLQSDFEYKLDREFLKGCKLSVTDDKNTVLALR NTLIESDLKYKEKHVKERGTCHAVPDTPQILLAKTVSNLVSENKYKDHVKKHLA QGSYTTLP TRDTPVHVKEVT KHVS DNTYK KKFVEKKGKSNYSIMLEPPEVKHAMEVAKKQSDVAYRKDAKEK LHYTTVADRPDIKKATQAAKQASEVEYRAKHKRKEGSHGLSMLGRPDIEMAKKAAKLSQVKYR ENFDKEKGKTPKYNPKDSQLYKVMKDANNLASEVKYKADLKKLHKPVTDMKESLIMNHVLNTS QCLASSYQKKKYEKS KGHYHTIPDNLEQLHLKEATELQSI VYKYKEKYEKERGKPMDFETPTY ITAKESQOMQSGKEYRKDY EESI KGRNLTGLEVT PALHLVKYATKIA SEKEYRKDLEESIRGK GLTEMEDTPDMLRAKNATQILNEKEYKRDLELEVKGRLNANANETPDFMRARNATDIASQIK YKQSAEMEKANFTSVVDTP EIIHAQQVKNLSSQKKYKEDA EKSMSYETVLDTPEIQRVRENO KNFSLQYQCDLKNKSGKITVVDTP EILRVKENQKNFSSVLYKEDVSPGTAIGKTPEMMRVK QTQDHISSVKYKEAIGQGTPIPDLP EVKRVKETQKHISVVMYKENLGTGIPITPEIERVKH NQENLSSVLYKENMGKGTPLPVTPEMERVKHNQENISSVLYKENMGKGTPLPVTPEMERVKHN QENISSVLYKENMGKGTPLPVTPEMERVKHNQENISSVLYKENMGKGTPLPVTPEMERVKHN ENFSSVLYKENLKGATPTPTPEMERVKRNQENISSVLYKENMRKATPTPTPEMERAKRNQE NISSVLYSDSFRKQIQGKAAYVLDTPEMRRVRETQRHISTVKYHEDFEKHKGCFTPVVTDPIIT ERVKKNMQD FSDIN YRGIQRKVEMEQRNDQDQETITGLRVWRTNPGSVFDYDPAEDNIQSR SLHMINVQAQRRSREQRSASALSVSGGEEKSEHSEAPDHHLSTYSDGGVFAVSTAYKHAKTT ELPQQRSSSVATQQTTVSSIPSHPSTAGKIFRMYDYMAADADEV SFKGDGAI INVQAIDEGW MYGTVQRTGRTGMLPANYVEAI		
	SEQ ID NO: 141	20881 bp	
NOV35b, CGI 19566-02 DNA Sequence	CCCACCTTTTGAGCAAGTT CAGCCTGGTTAAGTCCAAGCTGGTGATAAACTACAAAGCAGA ATACGAAGAAGACAGAGGCAAGGCTTCTCCCTCAGACCATAACTCAAGAAATATGGGGGTCT CGCAGTAATTTATGCTCTTTGCTTTTGTCTTTTCATAGTTTCTTGTATAGTTTGTCACTTA GGGCATCTCCTGCTGCCTTCAGCTGCCTTAAGCTGTGGGTCTCTGAAGCAGGAAGCATTAT AATCTGCTTTTCTTTTATTCTTTTCATAGTCTTCTTTGTATTTTGTCTGCTGAGGAAATTTATT TGGTAGATTGAAGGTTTGAACAGAGACTACAGAAACGAAAGAAAAGTCTGTATAAGCCAATG GTGTTGCGGAAGAAAATAACCCCATTCCTTGAAGTTTGTAGGTGCCACTACTACTCTGAAAAA TGGCAGATGACGAAGACTATAGGAGGTGGTGGAGTACTACACAGAAGAAGTGGTTTACGAAG AGGTGCCGGGAGAGACAATAACAAAATTTATGAGACTACGACAAACAAGGACATCTGACTATG AGCAATCAGAACTTCCAACAGCTCTGGCACAGCCAGCACTGGCACAGCCAGCATCAGCAA AGCCGGTGGAGAGGAGGAAGGTCATCCGGAAGAAAGTGGATCCTTCAAAGTTCATGACCCCTT ACATTGCACACAGTCAGAAAATGCAGGATCTTTTAGCCCAAATAAATAACAAGGAGAAGTTTG		

AGAAAACAAAAGGACAGCCATACGCCAGCACAAACAGATACTCCAGAACTTCGCAGAATCAAAA
AAGTACAAGATCAACTCAGTGAGGTTAAGTATCGAATGGATGGTGATGTTGCTAAGACTATAT
GTCACGTAGATGAAAAAGCAAAGGATATTGAACATGCAAAGAAAGTGTCCGACGCAAGTCAGTA
AGGTTTATACAAGCAGAACTGGGAAGACACCAAGGATAAGTACCTGCTTCCTCTGATGCC
CTGAACCTTGCCAGGCCGTTAAGAACACCGCCATGTTAGCAAGAAAGTGTACACTGAAGACT
GGGAAGCAGACAAAAGTTTGTGTTTACCCCTATAATGATAGCCCGAACTGAGGAGAGTTGCC
AGGCCCAGAAAGCTCTCAGTGATGTTGCCTACAAAAAGGTCTCGCTGAACAGCAAGCTCAAT
TCACGCCCTCTGGCCGATCCTCCAGATATAGAATTTGCCAAGAAAGTAACCAATCAAGTGAGCA
AGCAAAAATACAAAGAAGACTATGAAAAATAAATCAAAGGCAAATGGAGTGAGACACCTTGCT
TTGAAGTTGCAAATGCCAGAATGAATGCTGATAACATTAGCACAAGGAAATACCAGGAAGATT
TTGAAAACATGAAAGACCAGATCTACTTCATGCAGACCGAAACACCAGAGTATAAATGAATA
AAAAAGCTGGTGTGGCAGCTAGCAAGGTAAAAACAAAGAAGACTATGAAAAGAATAAAGGAA
AAGCAGATTATAATGTGCTTCCTGCTCAGAGAACCACAGCTTAGGCAGCTGAAGGCAGCAG
GAGATGCCCTAAGTGACAACTATACAAGGAAACTATGAAAAGACAAAGGCAAGAGCATAA
ATTACTGCGAGACCCCAAATTCAGCTCGATACTGTTCTGCAGAACTTCAGTAGTGATAAAA
AATATAAGATTCCCTACTTAAAGATATTTGGGACATTATGTAGGCAGCTTCGAGGATCCAT
ACCATTCACACTGCATGAAAGTCACAGCTCAAAACAGTGATAAAAACTACAAAGCAGAATACG
AAGAAGACAGAGGCAAGGCTTCTTCCTCAGACCATAACTCAAGAATATGAAGCAATTAAGA
AACTAGATCAGTGTAAGACCACACCTACAAAGTCCATCCAGATAAGACAAAATTCACCCAAG
TTACAGACTCTCCTGTTCTGCTACAGCCCAAGTCAATTCCAAACAACCTGAGTGACTTAAATT
ACAAAGCAAAACATGAAAGTGAAAGTTCAAGTGCCATATCCCCCTGATACTCCTG6T7TTA
TCCAGCACAAAGTCAATGCCTATACTTGAAGTGAATCTTTATAAGCAAGACTGGGAGAAGA
GCAAAGCCAAAAGTTTGACATTAAAGTGGATGCCATTCCCTGCTGGCAGCCAAAGCCAACA
CCAAGAACACCAGCGATGTGATGTACAAGAAAGACTATGAAAAAACAAAGGGAAAATGATTG
GAGTCCCTCAGCATTAATGACGATCCCAAGATGCTGCACTCCTGAAGGTGGCCAAAACCAGA
GTGATAGATTATACAAGGAAACTATGAGAAGACAAAGGCAAGAGTATGAATTACTGTGAGA
CCCCAAATATCAACTTGATACTCAGCTGAAGAACTTCAGTGAGGCTAGATATAAAGACTTAT
ATGTAAGGATGTTTGGGACATTATGTAGGCAGCATGGAGGACCCATATCACACACACTGCA
TGAAAGTTGCAGCTCAAAACAGTGATAAAGTTACAAAGCAGAATATGAAGAAGATAAAGGAA
AATGCTATTTCCCTCAGACAATAACACAAGAATATGACGCAATCAAGAAGCTGGACCACTGTA
AAGATCATACCTACAAAGTTCATCCAGATAAGACCAAATTCACGGCAGTCAGTATTCTCCTG
TACTGTTGCAAGCCAGCTCAACACGAAACAGCTTAGTGATCTGAATTACAAAGCAAAACATG
AAGGTGAGAGGTTCAAGTGCCATATACCAGCAGATGCTCCACAGTTTATCCAACACAGAGTCA
ATGCCCTATAATCTGAGTGATAATGTTTATAAGCAAGACTGGGAGAAGAGCAAGGCCAAGAAGT
TTGACATTAAAGTGGACGCCATTCCCTGTTGGCAGCCAAAGCCAACACCAAGAACACCAGCG
ATGTGATGTACAAGAAAGACTATGAAAAGAGCAAGGGAAAATGATTGGAGCCCTCAGCATTA
ATGACGATCCAAAGATGCTGCACCTCTTGAAGACAGCCAAAACAGAGTGATCGCGAATATC
GAAAAGATTATGAAAAGTCAAAACTATCTACACGGCACCTCTTGATATGCTCCAAGTCACTC
AAGCTAAGAAATCTCAGGCAATTGCCAGCGACGTTGATTATAAGCACATCTTACACAGTTACA
GCTACCCCTGATAGCATCAATGTGGACCTTGCCAAGAAGGCATATGCGCTGCAGAGCGATG
TTGAATACAAAGCTGACTACAATAGCTGGATGAAAGGTTGTGGCTGGGTGCCTTTTGGGTCT
TAGAAATGAAAAAGGCAAGCGAGCTTCAGACATCCTCAATGAGAAAAATATCGCCAAACATC
CAGACACCTCAAGTTTACCTCGATTGAAGATGCTCCAATTACAGTACAGTCTAAAATTAAACC
AGGCCCAGAGGAGTGATATCGCTTACAAAGCCAAAGGAGAGGAAATATTACAATTACAACC
TGCCACCAGACCTGCCCAAGTTCATCCAGGCTAAAGTTAATGCCTACAATATCAGTGAGAATA
TGTACAAAGCAGACTTGAAGACTTGAGCAAGAAGGGATATGACCTGAGAACTGATGCGATT
CCATCAGAGCTGCCAAAGCTGCCAGGCAAGCGGCGAGTGACGTTCACTACAAAAAGACTATG
AAAAGGCTAAAGGAAAAATGGTTGGCTTCAAAGTCTTCAAGATGACCTAAACTGGTTTCAAT
ATATGAACGTTGGCCAAGATACAATCAGATCGGGAGTATAAAAAAGACTATGAGAAGACAAAGT
CCAAATACAACACGCCCCATGATATGTTCAATGTGCTGGCGGCTAAGAAAGCCAGGATGTGG
TCAGCAATGTCAACTATAAGCATCTCTCCATCATACCTACTTGCTGACGCCATGGACC
TGGAGCTGTCTAAGAATATGATGCAGATACAGAGTGATAACGTTCTACAAGGAAGACTACAACA
ACTGGATGAAAGGCATTGGCTGGATTCTATTGGCAGTCTCGACGTCGAAAAAGTTAAAAAGG
CCGGTGATGCTCTGAATGAAAAGAAAGTACAGGCAACATCCAGACACCTCAAATTTACAGCA
TTGTGGACTCCCCAGTTATGGTCCAGGCAAAACAGAACACGAAGCAAGTCAAGTATATCTTAT
ACAAGGCTAAAGGAGAAGATGTGAAACATAAATACACCATGAGTCTCTGATCTTCTCAGTTT
TCCAGGCCAAGTGCAATGCTTACAGTATAAGTGACGCTGTTATAAACCAGGATTTGGCATGACT
TAATACGCAAGGCAACAATGTGCTGGGCGATGCTATTCCATCACTGACGCAAGGCATCGA
GAAACATTGCCAGTGATTATAAATACAAGGAAGCTTATGAGAAGTCAAAGGGAAAGCATGTGG
GTTTCAGAAGCCTCCAGGATGATCCCAAGCTGGTCCACTATATGAATGTGGCAAAGTGCAGT
CTGATCGTGAATACAAGAAGAACTATGAGAACACCAAAACAGCTACCATACCCCTGGGGACA
TGTTTACGATCACAGCTGCAAGATGGCCAGGATGTCGCTACCAATGTCAACTACAACACAGC
CATTGATCATTACACATACCTACCTGACGCCATGAGTCTTGAGCATAAGGAAATGTCAATC
AAATTCAGAGTGATAATGTGTATAAAGACGAGTATAACAGCTTCTTGAAGGGCATCGGATGGA
TCCCTATTGGTTCCCTGGAGGTGGAGAAAGTCAAGAAAGCAGGCGATGCATTAAATGAGAGGA
AGTATCGACAGCACCCAGATACCGTCAAGTTTACAAGTGTGCTGATTCCATGGGCATGATGT

	TGGCTCAGCATAACACAAAGCAGCTAAGTGATTGAACTACAAGGTAGAGGGAGAGAACTGA AGCACAAGTATACTATTGACCTGAATTGCCTCAGTTTATTCAAGCCAAAGTCAACGCCCTCA ACATGAGTGATGCTCATTATAAAGCAGATTGGAAGAAAACCATTCGCAAGGGCTATGATTGGA GACCAGATGCCATCCCAATTGTTGCTGCAAAAAGTTCAAGGAATATTGCTAGTGATTGCAAAAT ATAAGGAGGCCCTACGAGAAAGCCAAAGGCAAGCAAGTTGGATTCTCAGTCTTCAGGATGATC CTAAACTGGTTCACTACATGAATGTGGCCAAAATCCAGTCTGATCGTGAGTACAAAAGGGCT ATGAAGCCAGCAAGACCAAGTACCACACACCTCTGGATATGGTCAGTGTGACAGCTGCAAGA AATCTCAGGAGGTTGCCACCAACGCCAACTACAGACAGTCATACCACCCTACACTCTCCTGC CCGATGCCTTGAATGTGGAGCACTCCAGGAATGCCATGCAGATTCAGAGTGATAATCTGTACA AATCTGACTTCACCAATTGGATGAAAGGGATCGGCTGGGTGCCCATAGAGTCCCTGGAGGTGG AGAAGGCAAGAAAGCAGGAGAGATTCTTAGTGAGAAGAAGTATCGCCAGCACCCCGAGAAGC TGAAGTTCACCTACGCCATGGACACAATGGAACAGGCACTTAACAAGAGTAACAAACTGAACA TGGACAAGAGGCTCTACACTGAAAAATGGAACAAGGACAAGACCACCATTCATGTCATGCCTG ACACACCGGATATTTACTCTCCAGAGTAAACCAATCACCATGAGTGATAAACTGTACAAAG CTGGCTGGGAAGAGGAAAAGAAGAAAGGATATGACCTGAGGCCTGATGCCATTGCAATAAAGG CTGCAAGAGCCTCTAGAGACATTGCCAGTGATTACAAATACAAGAAAGCCTATGAACAGGCCA AAGGGAACACATTGGCTTCCGAGCCTGGAAGATGACCCCAAGCTGGTGCACTTCATGCAAG TGGCCAAAGATGCAGTCAGACCGGGAATACAAGAAGGGATATGAGAAATCCAAGACCTCCTTCC ACACCCCGGTGGACATGCTCAGTGTGGTGGCAGCCAAGAAGTCTCAGGAAGTGGCCACCAATG CCAACCTACAGGAACGTGATCCATACCTACAACATGCTTCCTGATGCCATGAGCTTTGAATTGG CCAAAAATATGATGCAGATTCAAAGTGATAATCAGTACAAGGCTGACTATGCTGACTTCATGA AGGGCATTGGATGGTCCCTCTGGGCTCCCTGGAAGCAGAGAAAAACAAGAAAGCCATGGAGA TTATTAGTGAAGAAGTACCGCAGCACCCAGACACTTTGAAGTATTCCACACTCATGGACT CGATGAACATGGTTTGGCCAGAATAATGCAAAAATATGAACGAACATCTCTACAAACAAG CATGGGAGGCTGACAAAACCAAAGTCCACATCATGCCTGATATCCCCCAGATTATTTTGGCAA AGGCAATGCAATTAATAAGTGATAAACTCTACAACTTTCCTTGAAGAGTCTAAAAAGA AAGGCTATGATCTCAGACCTGATGCAATTCTATCAAAGCTGCCAAGGCTTCCAGAGATATTG CAAGTGATTATAAATACAAGTACAATTATGAAAAAGGGAAGGGGAAAATGGTTGGTTTCCGCA GTCTCGAGGATGATCCCAATTAGTCCATTCATGCAAGTGGCTAAGATGCAATCTGATCGGG AGTACAAGAAAACTATGAGAACACAAAGACCAGCTACCACACCCCTGCCGACATGCTCAGTG TCACGGCTGCAAGGATGCCCAAGCCACATCACCACACTAACTACAAGCACCTGATTACACA AGTACATCCTCCTCCAGATGCAATGAACATTGAGCTGACCAGGAATATGAATCGCATACAGA GTGATAATGAATATAAGCAAGATTACAATGAATGGTACAAAGGGCTTGGCTGGAGTCCAGCAG GTTCTCTGGAAGTGGAGAAGGCCAAGAAAGCAACTGAATATGCCAGTGATCAGAAATACCGCC AGCACCCGAGCAACTTCCAGTTTAAAGAGCTGACTGATTCCATGGACATGGTGCTTGGCAAGC AGAATGCACATACCATGAACAAGCATTATACACCATTGATTGGAATAAAGATAAGACCAAGA TTCATGTGATGCTGATACACCAGATATTTACAAGCCAAGCAGAAATCAAACACTGTATAGTC AGAACTCTATAAACTTGGATGGGAAGAGCTTTGAAGAAAGGCTATGATCTCCAGTTGATG CAATTTCTGTACAGCTAGCTAAAGCTTCAAGAGACATTGCTAGTGATTATAAATACAAACAAG GCTACCGAAAGCAACTTGGCCACCATGTTGGATTCCGGAGTCTGCAAGATGACCCAAACTTG TGTTGTCCATGAATGTAGCCAAATGCAGAGTGAAGAGAATAACAAGAAGGACTTTGAGAAGT GGAAACTAAGTTCTCCAGCCAGTGGACATGTTGGGAGTGGTACTGGCCAAGAAGTGTGAGG AGTTGGTTAGTGACGTGGACTACAAGAACTACCTGCATCAGTGGACATGTCTGCCTGATCAGA ACGATGTTGTGAAGCTAAGAAAGTTTATGAAGTGAAGTGAGAATCTATATAAATCTGACC TTGAGTGGCTGAGAGGCATAGGATGGAGTCCCTTGGGTTCTTTAGAGGCAGAAAGACAAGC GGGCTTCGGAATCATCAGTGAGAAGAAATATCGTCAGCCTCCAGACAGAAACAAGTTCACCA GCATTCTGATGCCATGGATATAGTTCTGGCAAAGACAAATGCCAAAATAGGAGTGATAGAC TTTATAGAGAAGCTTGGGACAAGACAAGACTCAGATCCACATCATGCCTGATACACCTGACA TTGTTCTGGCTAAAGCAAATTAATCAACACAAGTGATAAACTCTACCGAATGGGTTATGAGG AGCTGAAGAGAAAAGGTTACGATCTTCTGTTGATGCCATACCAATCAAAGCAGCAAAAGCCT CCCGGGAATTCCAGTGAATACAAGTACAAGGAAGGCTTTCGCAAGCAGCTCGGCCACCACA TTGGTGGCCCGGAACATTGAAGATGACCCCAAGATGATGGGTCCATGATGTGGCCAAGATCC AGAGTGACAGGGAGTACAAGAAGGACTTTGAGAAGTGAAGACCAAGTTTCAAGCAGCCAGTGG ACATGCTGGGGTGGTGTGGCTTAAAGTGCCAGACCTTAGTCAGCGAGCTGGACTACAAGA ACTACCTGCACCAAGTGGACATGCCTGCCCGACCAGAGCGATGTATCCATGCTCGGCAGGCCT ATGACCTCCAGAGCGATAATTTGTACAAGTCAAGCTTCAAGTGGCTAAAAGGCATTGGCTGGA TGACTAGTGGTTCTCTCGAGGATGAGAAAAATAACAGAGCCACCCAGATTTTGTAGTGACCATG TTTACCGTCAGCACCCAGATCAATTAAGTTTTCAGCCTTATGGATTCCATACCAATGGTTT TGGCAAAAACAATGCTATTACCATGAATCATCGCTCTATACAGAAGCTTGGGATAAAGATA AAACCACTGTCCACATTATGCCAGATACCCCTGAAGTTTATTAGCTAAACAAAACAAGTAA ATTACAGTGAGAAATTGTATAAGCTTGGCCTAGAAGAAGCCAAGAGGAAGGTTATGACATGC GGGTAGATGCCATTCTATCAAGGCAGCCAAGGCTCCAGAGATATTGCAAGTGAATTCAGT ACAAAGAAGGCTATCGTAAGCAGCTCGGCCACCACATTGGTGCCGAGCTTACCTGATGAC CCAAGATGATGGTCCATGCACGTGGCCAAGATCCAGAGTGACAGGGAGTACAAGAAGGACT TTGAGAAGTGAAGACCAAGTTCAAGCAGCCAGTGGACATGCTGGGGTGGTGTGGCCAAGA AGTGCCAGACCTTAGTCAGCGATGTGGACTACAAGAACTACCTGCACCAAGTGGACATGCCTGC
--	---

CCGACCAGAGCGACGTCATCCATGCTCGGCAGGCCTATGACCTCCAGAGCGATAATATGTACA
AGTCTGATCTCCAGTGGATGAGAGGCATTGGCTGGGTGTCCATTGGCTCTTTGGATGTGGAAA
AATGCAAAAGGGCAACTGAAATTTTGTAGTGATAAAATCTATCGCCAGCCTCCAGACAGATTCA
AATTTACCAGTGTGACTGACTCTCTGGAACAAGTGTGGCCAAGAACAATGCTCTCAACATGA
ATAAGCGTTTATACACAGAGGCTGGGACAAAGACAAGACTCAAATTCACATAATGCTGATGATA
CACCAGAGATTATGTTGGCAAGGCAGAACAAAATCAACTACAGTGAGACTCTATACAACTTG
CCAATGAAGAAGCAAAAAAGAAAGGCTACGACTTGCGAAGTGACGCCATCCCCATCGTGGCTG
CCAAGGCCTCCAGGGACGTTATCAGTGATTACAAATACAAAGATGGTTACCGCAAGCAGCTCG
GCCACCACATTGGAGCCCGGAACATTGAAGATGACCCCAAGATGATGTGGTCCATGTCATGTGG
CCAAGATCCAGAGTGACAGGGAGTATAAGAAGGACTTTGAGAAGTGAAGACCAAGTTCAGCA
GCCCAGTGGACATGCTGGGAGTGGTGTAGCCAAGAAGTGCCAGACCTTAGTCAGCGATGTGG
ACTACAAGAACTACCTGCACGAGTGAGCGTGCCTGCCCGACCAAGATGATGTCATCCATGCTC
GGCAGGCCTATGACCTCCAGAGCGATAACATTTACAAATCTGATCTCCAGTGGCTGAGAGGCA
TTGGCTGGGTCCCCATTGGGTCTATGGATGTGGTCAAGTGCAAGAGAGCTGCTGAATACCTGA
GTGATAACATCTACCGCCAGCCTCCGACAAAGCTGAAATTTACCAGTGTGACTGACTCTCTAG
AGCAGGTGCTGGCCAAGAACAATGCTCTCAATATGAACAAGCGCTTATACACAGAAGCCTGGG
ACAAAGACAAGACCCAAGTCCATATTATGCCTGATACACCTGAAATCATGTTGGCAAGACAAA
ATAAAATAAATTATAGTGAGAGCCTCTATCGTCAGGCCATGGAAGAAGCCAAGAAAGAAGGCT
ATGACTTGAGAAGTGATGCCATTCCCATTGTGGCTGCCAAGGCCTCTCGGGATATTGCCAGTG
ATTACAAATACAAAGAAGCATATCGTAAGCAGTTGGGTCAACCATTTGGCGCCGAGCAGTAC
ACGATGACCCCAAGATAATGTGGTCCCTCCACATTGCCAAAGTGACAGTGACCGTGAGTACA
AGAAAGATTGTGAGAAATACAAGACAAGGTACAGCAGCCAGTGACATGCTTGGTATCGTTT
TGGCCAAGAAGTGTGACACCTTGGTCAGCGATGTGGACTATAAACATCCTCTGCATGAATGCA
TCTGCTGCCCGACCAAGATGACATCATTGATGCACGGAAGCCTATGACCTCCAGAGTGACA
ATTTGTATAAGTCAGACCTTGAATGGATGAAAGGCATTTGGCTGGGTTCCGATTGATTCCTTGG
AAGTTGTTAGGGCCAAGAGAGCTGGAGAATTACTTAGTGATACTATCTACCGTCAGCGTCCAG
AAACGCTGAAATTTACCAGTATAACGGACACTCCGGAGCAGGTGCTGGCAAAAAACAATGCTT
TAAACATGAATAAGCGCTTATATACTGAAGCCTGGGACAATGACAAGAAAATATTCTATGTCA
TGCCTGATACACCAGAAATCATGTTAGCCAAACTCAACCGAATAAATACAGTGATAAACTCT
ATAAACTTGCTTTGGAAGAGTCCAAGAAGGAAGGCTATGACTTGGCTGCTGGATGCCATTCCAA
TCCAAGCAGCCAAGGCTTCAAGAGATATTGCTAGTGATTACAAGTACAAGGAAGGCTACCGCA
AACAGCTTGGCCACCATATTGGGGCCCGAACATTAAAGGATGACCCGAAGATGATGTGGTCCA
TCCATGTGGCCAAGATCCAGAGTGACAGGGAGTACAAGAAGGAGTTTGAGAAGTGAAGACCA
AGTTCAGCAGCCAGTGGACATGCTGGGGGTGGTGTGGCCAAGAAGTGTGATCCTTGTA
GCGACATAGACTACAAGCATCCCTGTCATGAATGGACCTGCCTGCCTGATCAGAATGACGTC
TTCAGGCTCGGAAGGCCTATGACCTGCAGAGTGATGCTATTACAAATCTGATCTTGAGTGG
TGAGAGGCATAGGATGGGTTCCTATTGGCTCTGTAGAGGTGAGAAAGTGAAGAGAGCTGGAG
AAATCCTGAGTGACAGGAAGTATCGCCAGCTGCAGACCAGCTCAAATTCACATGCATTACCG
ACACTCCGGAAATGTCTTAGCAAGAATAATGCCTGACAATGAGCAAGCATTTATACACAG
AAGCTTGGATGTGACAAAACCTCCATCCAGTGATGCCAGACCCAGATATCTGTCTGG
CCAAGAGTAATCTGCCAATATCAGCCAAAACTTACACCAAGGGATGGGATGAATCAAAGA
TGAAGGACTATGATCTGAGAGCAGATGCTATTTCCATCAAAGTGCCAAGGCCTCCAGGGACA
TCGCCAGTGACTACAAATACAAGGAAGCCTATGAGAAACAGAAAGGCCACCATTTGGAGCCC
AGAGCATTGAAGATGATCCCAAGATTATGTGTGCCATACATGCAGAAAAAATCAAAGTGAAA
GGGAGTACAAGAAGGAATTCAAAAGTGGAAAACCAAGTTCTTAGCCCAAGTGACATGTTAA
GCATCTTGCTGGCCAAGAAATGTCAGACTTTGGTCACTGACATTTATTATCGCAATTACCTGC
ATGAATGGACATGCATGCCGATCAAAACGACATTATCCAAGCAAAAAAGGCCTATGACCTGC
AGAGTGATGCCCTCTACAAGGCTGACTTGGAGTGGTGTGGCTGGCATTGGCTGGATGCCCAAG
GGTCTCCTGAAGTGTGAGAGTCAAAAACGCCAGAATATCTTTGTGACAGTGTCTATCGGA
CGCCTGTGGTGAACCTTAAGTACACAAGCATTGTTGACACACCTGAAGTGGTCTTGTCTAAAT
CAAATGCTGAAAAATATTAGTATTCCAAAGTACAGAGAGGTTTGGGACAAGGATAAAACTTCAA
TACACATAATGCCAGATACTCCAGAAATTAATCTCGCTAGAGCAAATGCTCTTAATGTGAGCA
ATAAACTTTACCGTGAGGGCTGGGATGAAATGAAGGCGGGCTGTGATGTCGGCTGGATGCCA
TCCCCATCCAGGCTGCCAAGGCCTCCAGGGAGATTGCCAGTGACTATAAATATAAGCTTGACC
ATGAGAAGCAGAAGGGACACTACGTGGGCACCCTCACAGCCAGGGATGACAACAAGATCCGCT
GGGCCCTCATAGCTGACAAGCTCCAGAATGAACGAGAGTACCGGCTGGACTGGGCCAAATGGA
AGGCCAAGATCCAGAGCCTGTGGACATGCTTTCCATCTGCACTCTAAAAATTCCAGGCTC
TGGTCAGTGACATGGATTACCGCAATTACCTGCACCAAGTGGACCTGCATGCCCGACAGAAG
ATGTGATTACAGCCAAGAAGGCCTACGAAGTGCAGAGCGATAATGTTTACAAGGCTGACTTGG
AATGGTTGCGTGAATTTGGGTGGATGCCAAATGACTCCGTGTCCGTCAATCATGCCAAACATG
CCGCGGACATCTCAGTGAGAAAAATATCGCACAAAAATAGAACTCTCAACTTTACGCTG
TGGATGACAGAGTTGATTATGTGACAGCGAAACAAAGTGGCGAGATCCTCGATGATATTAAT
ACCGGAAAGACTGGAATGCCAACCAATCAAAGTACACCTCACAGAAACCCCTGCTGCACA
CTGCCCAGGAGGCTGCTAGGATACTGGACCAGTATCTCTACAAGGAAGGCTGGGAGAGACAAA
AAGCCACAGGTACATTTGCTCCAGATGCTGTGCCATTGTTTCATGCCCATCACTGCAATG
ACGTTCCAGAGTGAGCTGAAATACAAAGCTGAACATGTGAAGCAAAAGGTCATTATGTTGGT

TCCCACGATGAGAGATGATCCTAAGCTGGTTTGGTTTGAGCATGCAGGCCAGATTGAGAATG
AGAGACTATACAAAGAGGACTATCACAAAACAAAGGCCAAAATCAATATACCTGCTGATATGG
TGTCAGTCTTGGCCGCCAAGCAGGGGCAGACCCTTGTCAGTGATATTGATTATCGTAATTACT
TGCACCAATGGATGTGTCATCTGACCCAGAACGATGTTATTACAGGCAAGAAAGGCCTATGACC
TACAGAGTGATAATGTCTACAGAGCTGACCTGGAGTGGCTCCGAGGCATTGGCTGGATCCAC
TGGATTCTGTGGACCATGTAAGGGTTACTAAGAACCCAGGAAATGATGAGTCAGATCAAATATA
AGAAAAATGCCCTTGAAAATATCCTAACTTTACAAGTGTGGTGGATCCTCCAGAGATTGTTT
TAGCCAAGATTAATTCTGTCAATCAAAGTGATGTAAAATATAAAGAAACATTTAATAAAGCAA
AGGGCAAATATACGTTTTCCACAGATACACCACATATCTCCACTCCAAGACATGGGAAAC
TCTACAGTACTATACTGTATAAAGGGGCGTGGGAGGGCACCAAGGCCTATGGCTACACCCTGG
ATGAGCGCTACATTCCCATTGTTGGAGCCAAGCATGCTGATCTGGTGAACAGTGAGCTTAAAT
ACAAAGAGACATATGAGAAGCAGAAAGGCTACTACCTGGCTGAAAAGTGATCGGTGAATTCC
CTGGTGTGGTTCACTGTCTGGATTTCAAAAGATGAGGAGTGGTTGAACACAGAAAACATT
ATGAGGATACCAAAGCAAATGTTTATATCCCCAATGACATGATGAATCAGTGTGGCTAATA
GGTGCCAGTACATCCTCAGTGACCTGGAGTATCGACACTATTTCCACCAGTGGACGTCTCTTC
TGGGAAGAACCAATGTTATACGCGTCCGAAACGCCAGGAGATCTTGAGTGATAATGTGTATA
AAGATGACCTGAATTGGTTGAAAGGCATTGGTTGCTACGTTTGGGATACCCCCAAATCCTCC
ATGCCAAGAAATCATACGACCTTCAGAGTCAGTACAATATACAGCAGCAGGTAAAGAAAATC
TACAAAATATAATCTGGTCACAGACACGCCCTCTATGTGACTGCTGTTTACAGAGTGGCATT
ATGCCAGTGAGGTAAAATATAAAGAAAATATCATCAGATTAAAGACAAATACACAACAGTTC
TAGAAACAGTGGATTATGACAGAACCAGAAACCTGAAGAATCTTACAGCAGTAACCTGTACA
AGGAGGCTGGGATAGAGTGAAAGGCCACAGCTACATCTGCCTTCCAGACCTTGTCCCTGA
CACAGGCCAAGAACAGAGCATCTGGCCAGCCATATCAAATATCGGGAAGAAATATGAAAAGT
TCAAAGCTCTTTATACGTTACCAAGAAGTGTGACGATGATCCGAACACAGCAGGTGCCTCC
GAGTTGGCAAGCTTAACATCGATCGCCTGTACAGATCAGTTTATGAAAAGAACAAAGATGAAAA
TCCACATCGTGCCGACATGGTAGAGTGGTTACTGCCAAGGATCCCAGAAGAAAGTCAGTG
AGATTGATTACCGCTGCGCTCCACGAATGGATTGCCCACCCGACTTGCAGTCAZTGATC
ACGTCAGGAAAGTCACAGATCAGATCAGCGATATTGTATACAAGGATGACCTCAACTGGCTGA
AAGGCATTGGTTGCTACGTTCTGGGACACTCCTGAAATCCTCCATGCCAAGCATGCTTATGATC
TACGTGATGATATCAAGTATAAAGCTCACATGTTGAAAACAAGGAATGACTACAAGCTTGTC
CAGATACACCAGTCTACGTGACAGGCTGTCAAAGTGGGAAACAGTAAAGTACGCTGTCTACC
ACTATGACTATGTGCACAGTGTACAGGGCAAAGTGGCTCCAACTACAAAACCGTGGATCTGG
ACCGGGCCCTTCATGCATACAAGCTCCAGAGTTCGAATCTATACAAAACAGCCTGCGCACCC
TGCCCACTGGATATAGACTTCCAGGTGACACTCCTCACTTCAAACACATCAAGGACACCCGTT
ACATGAGCAGTTATTTCAAGTACAAAGAAGCCTATGAACACACCAAGGCATATGGGTATACAC
TTGGCCCCAAAGATGTTTCAATTGTCCAGTCCGGAGAGTCAACAAATGTTACCGAGGAGAC
TGTATCGGAATTGTACCACAACTGAAAGACAAGATCCATACAACCTCCCGATCCCCCTGAGA
TCCGCCAAGTCAAGAAGACACAAGAGGCTGTGAGTGAGTTGATCTACAAATCAGACTTCTTCA
AGATGCAGGGCCACATGATCTCTGCCATACACACCCCAAGTGATCCATTGCGCGTATGTGTG
GAGACATCACCAGTGATATTAATAACAAGAGGACTTGACAGGTCCTGAAGGATTTGGCTGGT
TCTGTATGACACTCCTGACATGGTCCGCTCCCGCACCTGCGGAAGCTCTGGTCTAATTACC
TATACACTGATAAGGCAAGGGAGATGCGAGACAAATACAAAGTGGTGCTTGACACTCCAGAAT
ACAGAAAAGTGAAGAAGTGAAGACACATCTGAGTGAGCTGGTCTACAGAGCTGCAGGCAAGA
AGCAGAAGTCAATCTTTACTTCACTTCTGATCTCTGATCTTTAAGAGCCAAGCGAGGGC
AGAAGCTTCAGAGTCAGTATCTGTATGTTGAACTTGCCACCAAGAGAGACCCCATCATCAGC
CTGGAAACCAGACCACAGCCTTGAAGCATGCTAAAGACGTGAAGGACATGGTCAGTGAGAAAA
AGTACAAGATTCAATATGAAAAGATGAAAGACAAGTACACTCCGTTCCAGATACGCCAATCC
TCATCAGAGCCAAGAGGGCTTACTGGAATGCCAGTGATCTACGCTACAAAGAAACATTTCAA
AGACCAAGGGAAATACCAACAGGTGAAAGATGCCCTAGACATTGTCTATCATCGCAAGTCA
CAGATGACATCAGTAAAAATAAATAACAAGGAGAACTACATGAGCCAGTTGGGTATCTGGAGGT
CCATTCTGATCGTCCAGAGCATTTCCACCACGAGCAGTCACTGACACAGTCAGTGATGTAA
AATATAAAGAAGACTTGACTTGGCTTAAAGGCATTGGTTGCTATGCTTATGATACCCCTGATT
TCACTCTGGCTGAAAAGAACAAGACTCTCTACAGCAAGTATAAGTATAAAGAAGTATTGAAA
GGACAAAGTCAGATTTCAAGTATGTTGCGGACTCTCCGATCAATAGGCATTTCAAGTGCAA
CTCAATTGATGAATGAGAAAAATACAGAGCTGATTATGAGCAGCGGAAGATAAATACCACC
TGGTAGTCGATGAGCCTAGACATCTGCTGGCTAAGACCCGACGACGACAGATCAGTCAGATCA
AATACAGGAAAAACTATGAAAAATCAAAGGACAAATTTACCTCAATTGTGGATCACTCCAGAAC
ACCTGCGTACTACAAAGTCAACAAACAAATCAGCGATACCTTTATAAATTGGAATCAACA
AGGCCAAACCCAGAGGCTACACCACAATCCACGACACGCCCATGTTGCTGCATGTCCGCAAGG
TTAAAGATGAAGTCAGTGATCTGAAATACAAAGAAGTATACCAAGAAATAAATCCAAGTCA
CCATTGAGCCAGATGCTGTTTATATCAAAGCAGCAAGGACGCTTACAAAGTCAACACCAATC
TGGACTATAAAGAACAGTACGAAGCCAACAAGCCCACTGGAAGTGGACTCTTGACCGACCGG
ACTTCTCCAGGCTGCAAGTCATCCCTGACGAAAGCGATTTTGAATATAAGCTGGACCGGG
AGTTCTCAAGGGTTGCAAGCTTTCTGTCACTGATGACAAAAACAGGTGCTCGCCCTCAGGA
ATACTTTAATGAAGATGATCTGAAATACAAAGAGAAACATGTCAAGGAAAGAGAACCTGCC
ATGCCGTACCTGACACGCCTCAGATCCTGCTGGCGAAGACTGTGACCAACCTGGTGTCTGAGA

	<p>ACAAGTACAAGGACCATGTCAAGAAGCACTTGGCACAGGGCTCATAACAACACTACCAGAGA CCCGGGACACTGTTACGTCGAAGGAAGTGACCAAGCATGTCAGTGATACAAATTACAAAAAGA AGTTTGTCAAGGAGAAAGGAAAATCCAACACTACTCCATCATGCTGGAGCCACCAGAGGTGAAAC ATGCTATGGAAGTGGCCAAGAAGCAAAGTGATGTCGCTTACAGAAAAGATGCCAAAGAGAACC TGCATTACACCACAGTGGCTGATCGACCAGACATCAAGAAGGCCACACAGGCAGCCAAACAGG CCAGTGAGGTGGAGTACAGAGCCAAGCACCAGCAAGGAAGGCAGCCATGGCTTAAGCATGCTCG GTCGCCCAGACATAGAAATGGCCAAGAAGGCAGCCAAGCTGAGCAGCCAGGTAAATACCGAG AAAATTTGATAAAGAAAAGGGCAAGACACCAAAATCAATCCAAAAGACAGCCAGCTCTACA AAGTCATGAAAGATGCTAATAATCTTGCAAGTGAGGTAAATACAAGGCTGACCTGAAGAAAC TTCACAAACCCGTGACTGACATGAAGGAGTCTCTGATCATGAATCATGCTCTGAATACAAGCC AACTTGCCAGTTCTTACCAGTACAAGAAGAAGTATGAGAAGAGTAAAGGCCACTACCACACCA TACCCGATAATCTGGAGCAGCTTACCTAAAAGAGGCCACAGAAATTACAGAGTATAGTGAAT ACAAGAAAAGTATGAAAAGGAACGAGGAAAACCCATGCTGGACTTTGAAAACCAACCGTACA TCACTGCCAAAGAGTCTCAGCAGATGCAGAGTGGGAAAGAATATAGGAAAGATTATGAAGAGT CCATTAAAGGCAGAAACCTGACTGGCTGGAGGTACGCCAGCTTTGTTACATGTCAAATATG CAACTAAAATAGCAAGCGAGAAAGAGTACAGGAAAGATCTAGAGGAAAGCATCCGTGGGAAGG GCCTCACTGAAATGGAAGATACACCTGACATGCTAAGAGCAAGAATGCCACTCAAATCCTCA ATGAGAAAGAATATAAGCGAGACCTGGAAGTCAAAGGAAGAGGCTGAATGCCATGG CCAATGAAACTCCGGATTTTATGAGGGCCAGGAATGCTACTGATATTGCCAGTCAGATTAAAGT ATAAGCAATCAGCAGAAATGGAGAAGCCAATTTCACTTCTGTGGTTGATACTCCAGAGATCA TTCATGCCCAACAGTCAAGAATCTTTCAAGCCAGAAAAGTACAAGGAAGATGCTGAGAAGT CCATGTCGTATTATGAGACTGTTTTGGACACCCAGAGATACAGAGAGTCCGGGAGAACCAA AGAAGTTTCAAGCTTCTCCAATACCAGTGTGACCTTAAAAACAGTAAAGGAAAATTACAGTTG TTCAAGACACGCCAGAAATCTGCGTGTAAGAAAATCAGAAGAATTCAGCTCGGTTTTAT ATAAAGAGGATGTCTACCAGGAACGGCTATCGGAAAGACACCTGAGATGATGAGAGTGAAC AAACACAGGACCACATTAGCTCGGTGAAGTATAAGGAAGCAATAGGACAAGGAACTCCAATCC CTGACCTGCCTGAAGTGAACGTGTGAAGGAGACGCAGAAGCACATTAGCTCGGTTATGTACA AAGAAAACCTGGGAACAGGCATTCCAACCACTGTGACTCCAGAGATTGAGAGAGTCAAAACGCA ATCAAGAGAACTTTAGCTCGGTTTTGTACAAAGAAAATTTGGGGAAGGAATCCCAACACCTA TCACTCCAGAGATGGAGAGAGTCAAACGCAATCAAGAGAATTTAGCTCGGTGTTATACAAAG AAAACATGGGCAAGGGAACCTTTTACCTGTCACTCCCGAGATGGAGCGAGTCAAAACAAATC AAGAAAATATTAGCTCGGTTTTGTACAAAGAAAATGTGGGGAAGCCACCGCAACCCCTGTCA CTCCTGAGATGCAGAGAGTCAAACGCAATCAAGAAAACATTAGCTCGGTGTTATACAAAGAGA ACCTGGGGAAGCAACCCCCACACCTTTTACTCCTGAGATGGAAAGAGTGAACGCAATCAAG AAAACTTTAGCTCGGTATTGTACAAAGAGAACATGAGAAAAGCAACTCCGACACCTGTTACTC CAGAGATGGAGAGAGCTAAGCGCAACCAAGAAAACATTAGCTCGGTTCTTTATCTGATAGTT TCCGGAACAAATACAAGGCAAGAGTGCCTATGTATTGGATACCCCGAGATGAGACGGGTGA GGGAGACCCAACGGCACATCTCAACGGTGAATATCATGAAGACTTTGAGAAACACAAGGGTT GCTTCACACCAAGTGGTGACAGATCTATCACTGAACGAGTAAAGAAGAACATGCAGGACTTCA GTGACATTAACACCGAGGTATTAGAGGAAAGTGGTAGAAATGGAACAAAAACGGAATGACC AAGATCAGGAGACTATTACAGGTTTACGTGTCTGGCGTACTAATCCTGGTTCCGTTTTTGACT ATGATCCAGCAGAGACAACATCCAGTCCCGAAGCTTACACATGATTAATGTCCAAGCTCAGC GCCGAGCCGGGAGCAGTCAGATCTGCCAGTGCCTAAGCGTCAGTGGGGGTGAGGAGAAGT CTGAGCATTCAAGAGCACCAGACCACCTTTTCACTTACAGCGACGGGGGTGCTTTGTCAG TCTCAACAGCTTACAAACATGCAAAACCAAGAGTCTCCACAACAACGATCATCTTCAGTTG CTACCAACAGACAACGGTATCTTCCATCCCATCTCATCCATCTACTGCTGGAATAATCTTCC GTGCCATGTATGACTATATGGCTGCTGATGCAGATGAGGTGTCCTTCAAGGATGGAGATGCCA TCATAAATGTTCAAGCAATTGATGAAGGCTGGATGTATGGCACTGTGCAGAGGACTGGCAGGA CCGGAATGCTCCAGCCAACCTACGTTGAAGCTATTAGGCATTTCAAAGCATCACACTTGCTCT GCAGGACTTACAGATCCTGCAGTCAATGTTTCGGTTTAGACTCTCCACTGTTACCTAAGTTCT CAAGCTGCCTATGGTTTTCTGTGTCAATGTGATTTATGGTAGTACCATCCTTTCTCCTTTGG GTTTTAAATAAGTTGCAGAACAGACACTTTAAAAGCTTCTGCAATATTATTCTGTGCCTAG AGTCTTTCTCCATTATAAATCATGTTTTAACATTATTCTTTCTAAAACAGGGATTTTGAATA TGCCAAAACATTAAAGGAAAAATAGCAGAGATGTTACCTTTTCTTGTGATTGCTAATGC TTATTATTCTAATTCAGTTCTGAAGTTATAAACTTATAATCAATACAAACCAGCACTAATA AAACCTCTAATTCTGCAAAAAA</p>
	<p>ORF Start: ATG at 441</p>
	<p>ORF Stop: TAG at 20448</p>
	<p>SEQ ID NO: 142</p>
	<p>6669 aa</p>
	<p>MW at kD</p>
<p>NOV35b, CG119566-02 Protein Sequence</p>	<p>MADDEDYEEVVEYYTEEVVEYVGETITKIYETTTTRTSDEYQSETSKPALAQPALAQPASA KPVERRKVIKKVDPSKFMTPYIAHSQKMDLFSFNKYKEKFEKTKGQPYASTTDTPELRRIK KVQDQLSEVKYRMDGDVAKTICHVDEKAKDIEHAKKVSQQVSKVLYKQNWEDTKDKYLLPDA PELVQAVKNTAMFSKKLYTEDWEADKSLFYPYNDSPELRRVAQAQKALSDVAYKKGLAEQQAQ FTPLADPPDIEFAKKVTNQVSKQYKEDYENKIKGKWSSETPCFEVANARMNADNISTRKYQED FENMKDQIYFMQTEPEYKMNKAGVAASKVKYKEDYENKNGKADYNVLPASENPQLRQLKAA GDALSDKLYKENYEKTKAKSINYCETPKFKLDTVLQNFSSDKKYKDSYKLDILGHYVGSFEDP</p>

YHSHCMKVTAQNSDKNYKAEYEEDRGKGFPPQITITQEYEAIKKLDQCKDHTYKVHPDKTKFTQ
 VTDSPVLLQAQVNSKQLSDLNKAKHESEKFKCHIPPDTAFIQHKVNAYNLSNLYKQDWEK
 SKAKKFDIKVDAIPLLAAKANTKNTSDVMYKKDYKNGKMGVLSINDDPKMLHSLKVAKNQ
 SDRLYKENYEKTKAKSMNYCETPKYQLDLTKNFSEARYKDLVYKDVLGHYVGSMPEDPYHTHC
 MKVAAQNSDKSYKAEYEEDKGCYFPQITITQEYDAIKKLDQCKDHTYKVHPDKTKFTAVTDS
 VLLQAQNTKQLSDLNKAKHEGERFKCHIPADAPQFIQHRVNAYNLSNLYKQDWEKSKAKK
 FDIKVDIPLLAAKANTKNTSDVMYKKDYKSKGKMIGALSINDDPKMLHSLKTAQNQSDREY
 RKDYKSKTIYTAPLDMQLQVTAQKKSQAIASDVYKHILHSYSYPPDSINVDLAKKAYALQSD
 VEYKADYNSWMKGCWVPFSGLEMEKAKRASDILNEKKYRQHPDTLKFTSIEDAPITVQSKIN
 QAQRSDIAYKAKGEEIHNYNLPDLPQFIQAKVNAYNISNMYKADLKDLSKKGVDLRTDAI
 PIRAAKAAQAAQSDVYKDYKAKGKMVGQSLQDDPKLVHYMNVAKIQSDREYKDYKTK
 SKYNTPHDMFNVAAKAQQVSVNMYKSHLHYTYLPDAMDLELSKNMMQIQSDNMYKEDYN
 NWMKGIGWPIGSLDVEKVKAGDALNEKKYRQHPDTLKFTSIVDSPVMVQAKQNTKQVSDIL
 YKAKGEDVKHYTMSPDLPQFLQAKCNAYSISDVCKYKRDWHLIRKGNVLDGAIPIAAKAS
 RNIAADYKKEAYEKSKGKHVGFSLQDDPKLVHYMNVAKLQSDREYKKNYENTKTSYHTPGD
 MVTITAAKMAQDVATNVNYKQPLHHYTYLPDAMSLEHTRNVNQIQSDNMYKDEYNSFLKGIGW
 IPIGSLVEKVKAGDALNEKKYRQHPDTLKFTSIVDSPMSGMMLAQHNTKQLSDLNKVEGEKL
 KHKYTIDPELPQFIQAKVNALNMSDAHYKADWKKTIKGYDLRPAIPIVAAKSSRNIAADCK
 YKAEYKAKGKQVGLSLQDDPKLVHYMNVAKIQSDREYKKGYEASKTYHTPLDMVSVTAAK
 KSQEVATNANYRQSYHHYTLPLDALNVEHSRNAMQIQSDNLYKSDFTNWMKGIGWVPIESLEV
 EKAKKAGEILSEKKYRQHPDKLFTYAMDTMEQALNKSNNKLNMDKRLYTEKWNKDKTTHVMP
 DTPDILLSRVNIQITMSDKLYKAGWEEKKKGYDLRPAIAIKAAASRDIAADYKKEAYEQ
 KGHIGFRSLEDDPKLVHFMQVAKMQSDREYKKGYEKSKTSFHTPVDMLSVVAAKKSQEVATN
 ANYRNVITHTNMLPDAMSFELAKNMMQIQSDNMYKADYADFMKGIGWLPGLSLEAKNKAME
 IISEKKYRQHPDTLKYSTLMDSMNMLVAQNNAKIMNEHLYKQAWADKTIVHMPDIPIQILA
 KANAINISDKLYKLSLEESKKGYDLRPAIPIKAAASRDIAADYKKNYKKGKGMVGFR
 SLEDDPKLVHSMQVAKMQSDREYKKNYENTKTSYHTPADMLSVTAADKADANITNTNYKHLIH
 KYILLPDAMNIELTRNMNRIQSDNEYKQDYNEWYKGLGWSVAGSLEVEKAKKATEYASDQKYR
 QHPSNFQFKKLTDSMDMVLAKQNAHTMKNHLYTIDWNKDKTTHVMPDTPDILQAKQNTLYS
 QKLYKLGWEEALKKGYDLVDAISVQLAKASRDIAADYKQGYRQKQLGHHVGFSLQDDPKL
 VLSMNVAKMQSEYKKEFKWTKFSSPVDMLGVVLAACKQELVSDVDYKNYLHQWTCPLDQ
 NDVVQAKVYELQSENLYKSDLEWLRGIGWSPGLSLEAKNKRAEIISEKKYRQPPDRNKFT
 SIPDAMDIVLAKTNAKNSDRLYREAWDKDKTQIHIMPDTPDIVLAKANLINTSDKLYRMGYE
 ELKRKGYDLVDAIPIKAAASREIASEYKKEGFRKQLGHHIGARNIEDDPKMMWSMHVAKI
 QSDREYKKEFKWTKFSSPVDMLGVVLAACKQTLVSDVDYKNYLHQWTCPLDQSDVIHARQ
 YDLQSDNLYKSDQLQWLKIGWMTSGSLEDEKNKRATQILSDHVYRQHPDQFKFSSLMDSIPMV
 LAKNNAITMNRHLYTEAWDKDKTTHVIMPDTPEVLLAKQNKVNYSEKLYKLGLEAKRKGDM
 RVDIPIKAAASRDIAEFKYKGYRQKQLGHHIGARAI RDPKMMWSMHVAKIQSDREYKKE
 FEKWKTKFSSPVDMLGVVLAACKQTLVSDVDYKNYLHQWTCPLDQSDVIHARQAYDLQSDNMY
 KSDLQWMRGIGWVSIGSLDVEKCRATEILSDKIYRQPPDRFKFTSVTDSLEQVLAKNNALNM
 NKRLYTEAWDKDKTQIHIMPDTPEIMLARQNKINYSYLYLANEAKKGYDLRSDAIPIVA
 AKASRDVSDYKKGYRQKQLGHHIGARNIEDDPKMMWSMHVAKIQSDREYKKEFKWTKF
 SPVDMLGVVLAACKQTLVSDVDYKNYLHEWTCLPDQNDVIHARQAYDLQSDNLYKSDQLWLRG
 IGWVPIGSMVDVVKRAAEILSDNIYRQPPDKLFTSVTDSLEQVLAKNNALNMNRHLYTEAW
 DKDKTQVHIMPDTPEIMLARQNKINYSYLYRQAMEEAKKGYDLRSDAIPIVAAKASRDIA
 DYKKEAYRQKQLGHHIGARAVHDDPKIMWSLHIAKVQSDREYKKEFKYTRYSSPVDMLGIV
 LAKKQTLVSDVDYKHLHECICLPDQNDIIHARKAYDLQSDNLYKSDLEWMMKGIGWVPIDSL
 EVVRKRAAGELSDTIYRQRPETLKFTSITDTPQVLAKNNALNMNRHLYTEAWDNKKTTHV
 MPDTPPEIMLAKLNRINYSKLYKLALESKEGVDLRDAIPIQAAKASRDIAADYKKEGYR
 KQLGHHIGARNIKDDPKMMWSIHVAKIQSDREYKKEFKWTKFSSPVDMLGVVLAACKQILV
 SDIDYKHLHEWTCLPDQNDVIQARKAYDLQSDAIYKSDLEWLRGIGWVPIGSEVEKVKRAG
 EILSDRKYRQPADQLKFTCITDTPPEIVLAKNNALTMSKHYLATEAWADKTSIHVMPDTPDILL
 AKSNSANISQKLYTKGWDESKMDYDLRADAISIKSAAKASRDIAADYKKEAYEKQKGGHIGA
 QSIEDDPKIMCAIHAEKIQSERYKKEFKWTKFSSPVDMLSILLAKKQTLVTDIYRNYL
 HEWTCMPDQNDIIQAKKAYDLQSDALYKADLEWLRGIGWMPQGSPEVLRVKAQNIFCDSVYR
 TPVVNLKYTSIVDTPEVVLAKSNAENISIPKYREVWDKDKTSHIMPDTPEINLARANALNVS
 NKLYREGWDEMAGCDVRLDAIPIQAAKASREIASEYKKEKLDHEKQKGYHVGTLTARDNKRIR
 WALIADKLQNEREYRLDWAQKAKIQSPVDMLSILHSKNSQALVSDMDYRNYLHQWTCMPDQ
 DVIQAKKAYELQSDNMYKADLEWLRGIGWMPNDSVSVNHAKHAADI FSEKKYRTKIETLNFTP
 VDDRVDYVTAQSGEILDDIKYRKDNATKSKYTLTETPLHTAQEAARILDQYLYKEGWERQ
 KATGYILPPDAVFPVHAHHCNDVQSELKYKAEHVQKQGHYVGVPTMRDPPKLVVFEHAGQIQN
 ERLYKEDYHKTAKINIPADMVSVLAQKQGTLSIDIDYRNYLHQWMCHPDQNDVIQARKAYD
 LQSDNMYRADLEWLRGIGWIPLDSVDHVRVTKNQEMMSQIKYKNALENYPTSVVDPPEIV
 LAKINSVNQSDVYKETFNKAAGKYTFSPDTPHISHSKDMGKLYSTILYKGAWEGTKAYGYTL
 DERYIPIVGAHADLVNSELKYKETYKQKGYHLAGKVI GEPGVVHCLDFQKMSALNRYKH
 YEDTKANVHIPNDMMNHVLAQRQYILSDLEYRHYFHQWTSLEENPVIRVRNAQEILSDNMY

	<p>KDDLNLWKGIGCYVWDTPQILHAKKSYDLQSQLQYTAAGKENLQNYNLVTDTPLYVTAVQSGI NASEVKYKENYHQIKDKYTTVLETVDYDRTRNLKNLYSSNLYKEAWDRVKATSILPSSTLSL THAKNQKHLASHIKYREEYEKFKALYTLPRSVDDDPNTARCLRVGKLNLDRLYRSVYEKNMK IHIVPDMVEMVTAKDSQKKVSEIDYRLRLHEWICHPLQVNDHVRKVTDQISDIVYKDDLNLW KGIGCYVWDTPQILHAKHAYDLRDDIKYKAHMLKTRNDYKLVTDTPVYVQAVKSGQLSDAVY HYDYVHSVRGKVAPTTKTVLDRALHAYKLQSSNLYKTSRLTLPTGYRLPGDTPHFKHIDTR YMSSYFKYKEAYEHTKAYGYTLGPKDVPFVHVRVNVTSERLYRELYHKLKDKIHITTPDPPE IRQVKKTQEAUSELIYKSDFFKMQGHMISLPYTPQVIHCYVGDITSIDIKYKEDLVKLGFGC FLYDTPDMVRSRHLRKLWSNYLYTDKAREMRDKYKVLDTPYRKVQELKTHLSELVYRAAGK KQKSIFTSVDPDTPDLLRAKRGQKLSQYLYVELATKERPHHAGNQTALKHAKDVKDMVSEK KYKIQYEKMKDKYTPVPDTPILIRAKRAYWNASDLRYKETFOKTKGKYHTVKDALDIVYHRKV TDDISKIKYKENYMSQLGIWRSIPDRPEHFHRAVTDVSDVYKEDLTWLKIGCYAYDTPD FTLAEKNKTLYSKYKYKEVFERTKSDFKYVADSPINRHFKYATQLMNEKKYRADYEQRKDKYH LVVDEPRHLLAKTRSDQISQIKYRKNYEKSDDKFTSIVDTPHLRRTTKVKNQISDILYKLEYN KAKPRGYTTIHDTMPLLHVRKVKDEVSDLKYKEVYQRNKSNCETIEPDAVHIKAAKDAYKVNTN LDYKQYEAANKAHWKWTPDRPDFLQAAKSSSQSDFEYKLDREFLKGCKLSVTDDKNTVLALR NTLIESDLKYKEKHVKERGTCHAVPDPDQILLAKTVSNLVSENKYKDHVKHKLAAQGSYTTLP TRDTVHVKEVTKHVSNTNYKKKFVKEKGKSNYSIMLEPPEVKHAMEVAKQSDVAYRKDAKEN LHYTTVADRPDIKKATQAAKQASEVEYRAKHKRKEGSHGLSMLGRPDIEMAKKAALSSQVKYR ENFDKEKGKTPKYNPKDSQLYKVMKDANNLASEVKYKADLKLHKPVTDMKESLIMNHVNTS QLASSYQYKKKYEKSQGHYHTIPDNLEQLHLKEATELQSIKYKEKYEKERGKPMDFETPTY ITAKESQMQSGKEYRKDYESI KGRNLTGLEVT PALHVKYATKIASEKEYRKDEESIRGK GLTEMEDTPDMLRAKNATQILNEKEYKRDLELEVKGRLNANANETPDMRARNATDIASQIK YKQSAEMEKANFTSVVDTPQILHAKKSYDLQSQLQYTAAGKENLQNYNLVTDTPLYVTAVQSGI KNFSLQLQYQCDLKNKSGKITVVDTPQILRVKENQKNFSSVLYKEDVSPGTAIGKTPEMMRVK QTQDHISVVKYKEAIGQGTPIPDLPVVKRVKETQKHISVVMYKENLGTGIPPTVTPPIERVKR NQENFSSVLYKENLKGIPPTPIPEMERVKRNQENFSSVLYKENMGKGTPLPVTPEMERVKRN QENISSVLYKENVGKATATPVTPEMRVKRNQENISSVLYKENLGKATPTPTPEMERVKRNQ ENFSSVLYKENMRKATPTPTPEMERAKRNQENISSVLYSDSFRKQIQGKAAYVLDTPEMRRV RETQRHISTVKYHEDFEKHKGCTFPVVDTPITERVKKNMQDFSDINRYGIRKRVVEMEQKRN QDQETITGLRVWRTNPGSVFDYDPAEDNIQSRSLHMINVQAQRRSREQRSASALSVSQGEK SEHSEAPDHHLSTYSDGGVFAVSTAYKHAKTTELPPQRRSSSVATQQTTSVSSIPSHPSTAGKIF RAMYDYMAADADEVSKDGDIAINVQAIDEGWMYGTVQRTGRTGMLPANYVEAI</p>
	<p>SEQ ID NO: 143 20194 bp</p>
NOV35c, CGI19566-03 DNA Sequence	<p>CCACTACTACTCTGAAAAATGGCAGATGACGAAGACTATGAGGAGGTGGTGGAGTACTACACA GAAGAAGTGGTTTACGAAGAGGTGCCGGGAGAGACAATAACAAAAATTTATGAGACTACGACA ACAAGGACATCTGACTATGAGCAATCAGAACTTCCAAACAGGCTCTGGCAGAGCCAGCACTG GCACAGCCAGCATCAGCAAAGCCGGTGGAGAGGAGGAAGGTCATCCGGAAGAAAGTGGATCCT TCAAAGTTCATGACCCCTACATTGACACAGTCAGAAAATGCAGGATCTTTTAGCCCAAAT AAATACAAGGAGAAGTTTGAGAAAAAAAAGGACAGCCATACGCCAGCACAACAGATACTCCA GAACCTCGCAGAAATCAAAAAAGTACAAGATCAACTCAGTGAGGTTAAGTATCGAATGGATGGT GATGTTGCTAAGACTATATGTCACGTAGATGAAAAAGCAAAGGATATTGAACATGCAAGAAA GTGTCGACAGCAAGTCAGTAAGGTTTATACAAGCAGAACTGGGAAGACACCAAGGATAAGTAC CTGCTTCTCTGATGCCCTGAACTTGTCCAGGCCGTTAAGAACACCGCCATGTTACGCAAG AACTGTACACTGAAGACTGGGAAGCAGACAAAAGTTTGTCTTACCCCTATAATGATAGCCCG GAACTGAGGAGAGTTGCCAGGCCAGAAAAGCTCTCAGTGATGTTGCCTACAAAAAAGGTCTC GCTGAACAGCAAGCTCAATTCACGCCTCTGGCCGATCTCCAGATATAGAATTTGCCAAGAAA GTAACCAATCAAGTGAGCAAGCAAGCAAAAATACAAAGAAGACTATGAAAATAAATCAAAGGCAAA TGGAGTGAGACACCTTGCTTTGAAGTTGCAATGCCAGAAATGAATGCTGATAACATTAGCACA AGGAAATACCAGGAAGATTTTGAACAATGAAAGACCAGATCTACTTCATGCAGACCGAAACA CCAGAGTATAAAATGAATAAAAAAGCTGGTGTGGCAGCTAGCAAGGTAATAACAAAGAAGAC TATGAAAAGAATAAAGGAAAAGCAGATTATAATGTGCTTCTGCTTCAGAGAACCACAGCTT AGGCAGCTGAAGGCAGCAGGAGATGCCCTAAGTGACAACTATACAAGGAAAACATGAAAAAG ACAAAAGCAAAGAGCATAAATTACTGCGAGACCCCAAAATCAAGCTCGATACTGTTCTGCAG AACTTCAGTAGTGATAAAAAATATAAGATTCTACTTAAAGATATTTGGGACATTATGTA GGCAGCTTCGAGGATCCATACCATTCACACTGCATGAAAGTCACAGCTCAAAACAGTGATAAA AAGACAAAATTCACCCAGTTACAGACTCTCCTGTTCTGCTACAAGCCCAAGTCAATCCAAA CAACTGAGTGACTTAAATTACAAAGCAAAACATGAAAGTGAAAGTTCAAGTGCCATATCCCC CTGTATCTCTGCTTTTATCCAGCACAAAGTCAATGCCTATAACTTGAGTGATAATCTTTAT AAGCAAGACTGGGAGAAGAGCAAGGCCAAAAGTTTGACATTAAGTGGATGCCATTCCTCTG CTGGCAGCCAAAGCCAAACCAAGAACACCAGCGATGTGATGTACAAGAAAGACTATGAAAAA AACAAAGGAAAATGATTGGAGTCTCAGCATTAATGACGATCCCAAGATGCTGCACTCCTTG AAGGTGGCCAAAACAGAGTGATAGATTATACAAGGAAAACATGAGAAGACAAAGGCAAG</p>

	AGTATGAATTACTGTGAGACCCCAAAATATCAACTTGATACTCAGCTGAAGAACTTCAGTGAG GCTAGATATAAAGACTTATATGTAAAGGATGTTTTGGGACATTATGTAGGCAGCATGGAGGAC CCATATCACACACTGCATGAAAGTTGCAGCTCAAAACAGTGATAAAAGTTACAAAGCAGAA TATGAAGAAGATAAAGGAAAATGCTATTTCCCTCAGACAATAACACAAGAAATATGACGCAATC AAGAAGCTGGACCAGTGTAAGATCATACCTACAAAGTTTCATCCAGATAAGACCAAAATTCAG GCAGTCACTGATTCTCCTGTACTGTTGCAAGCCAGCTCAACACGAAACAGCTTAGTGATCTG AATTACAAAGCAAAACATGAAGGTGAGAGGTTCAAGTGCCATATACCAGCAGATGCTCCACAG TTTATCCAACACAGAGTCAATGCCTATAATCTGAGTGATAATGTTTATAAGCAAGACTGGGAG AAGAGCAAAGCCAAGAAGTTTGACATTAAAGTGGACGCCATCCCCCTGTTGGCAGCCAAAGCC AACACCAAGAACACCAGCGATGTGATGTACAAGAAAGACTATGAAAAGAGCAAAGGGAAAATG ATTGGAGCCCTCAGCATTATGACGATCCAAAGATGCTGCACCTCTGAAGACAGCCAAAAAC CAGAGTGATCGCGAATATCGAAAAGATTATGAAAAGTCAAAAATCTACACGGCACCTCTT GATATGCTCCAAGTCACTCAAGCTAAGAAATCTCAGGCAATTGCCAGCGACGTTGATTATAAG CACATCTTACACAGTTACAGTACCCCTGATAGCATCAATGTGGACCTTGCCAAAGGAGCA TATGCGCTGCAGAGCGATGTTGAATACAAAGCTGACTACAATAGCTGGATGAAAGGTTGTGGC TGGGTGCCCTTTGGGTCTTAGAAATGGAAAAGGCAAGCGAGCTTCAGACATCCTCAATGAG AAAAATATCGCCAACATCCAGACACCTCAAGTTTACCTCGATTGAAGATGCTCCAATTACA GTACAGTCTAAAATTAACCAGGCCAGAGGAGTGATATCGCTTACAAAGCCAAAGGAGAGGAA ATTATTACAATTACAACCTGCCACCAGACCTGCCCCAGTTTCATCCAGGCTAAAGTTAATGCC TACAATATCAGTGAGAATATGTACAAAGCAGACTTGAAAGACTTGAGCAAGAAGGGATATGAC CTGAGAACTGATGCGATTCCCATCAGAGCTGCCAAAGCTGCCAGGCAGGCGGCGAGTGACGTT CAGTACAAAAAGACTATGAAAAGGCTAAAGGGAAAATGGTTGGCTTCCAAGTCTTCAAGAT GACCTTAACTGGTTTCAATTATATGAACGTGGCCAAAGATACAATCAGATCGGGAGTATAAAAA GACTATGAGAAGACAAAGTCCAATACAACACGCCCCATGATATGTTCAATGTCGTGGCGGCT AAGAAAGCCAGGATGTGGTGCAGCAATGTCAACTATAAGCATCTCTCCATCATTACACCTAC TTGCCTGACGCCATGGACCTGGAGCTGTCTAAGAACATGATGCAGATACAGAGTGATAACGTC TACAAGGAAGACTACAACAACCTGGATGAAAGGCATTGGCTGGATTCTTATGGCAGTCTCGAC GTGAAAAAGTTAAAAAGGCCGGTGATGCTCTGAATGAAAAGAGTACAGGCAACATCCAGAC ACCCTCAAATTTACAGCATTTGTGGACTCCCCAGTTATGGTCCAGGCAAAACAGAACAGGAAG CAAGTCAGTGATATCTTATACAAGGCTAAAGGAGAAGATGTGAAACATAAATACACCATGAGT CCTGATCTTCTCAGTTTCTCAGGCCAAGTGCAATGCTTACAGTATAAGTACGTCGTGTTAT AAACGGGATTTGGCATGACTTAATACGCAAGGGCAACAATGTGCTGGGCGATGCTATTCCCATC ACTGCAGCCAAGGCATCGAGAAACATTGCCAGTGATTATAAATACAAGGAAGCTTATGAGAAG TCAAAGGGAAAGCATGTGGGTTTCAAGAGCTCCAGGATGATCCCAAGCTGGTCCACTATATG AATGTGGCAAAGCTGCAGTCTGATCGTGAATACAAGAAGAACTATGAGAACCAAAAACCAGC TACCATACCCTGGGGACATGGTTACGATCACAGCTGCAAAGATGGCCAGGATGTCGCTACC AATGTCAACTACAAACAGCCATTGCATCATACACATACCTACCTGACGCCATGAGTCTTGAG CATACGAGGAATGTCAATCAAATCAGAGTGATAATGTGTATAAAGACGAGTATAACAGCTTC TTGAAGGGCATCGGATGGATCCCTATTGGTTCCCTGGAGGTGGAGAAGGTCAAGAAAGCAGGC GATGCATTAAATGAGAGGAAGTATCGACAGCACCCAGATACCGTCAAGTTCACAAGTGTGCCT GATTCCATGGGCATGATGTTGGCTCAGCATAACACAAAGCAGCTAAGTGATTGAACTACAAG GTAGAGGGAGAGAACTGAAGCACAAGTATACTATTGACCCTGAATTGCCTCAGTTTATCAA GCCAAAGTCAACGCCCTCAACATGAGTGATGCTCATTATAAAGCAGATTGGAGAGAAACCATT CGCAAGGGCTATGATTGAGACCAGATGCCATCCCAATTGTTGCTGCAAAAGTTCAAGGAAT ATTGCTAGTGATTGCAAAATATAAGGAGGCTACGAGAAAGCCAAAGGCAAGCAAGTTGGATT CTCAGTCTTCAGGATGATCCTAACTGGTTCACTACATGAATGTGGCCAAAATCCAGTCTGAT CGTGAGTACAAAAGGGCTATGAAGCCAGCAAGACCAAGTACCACACACCTCTGGATATGGTC AGTGTGACAGCTGCAAGAAATCTCAGGAGGTTGCCACCAACGCCAACTACAGACAGTCATAC CACCCTACACTCTCTGCCCCGATGCCCTGAATGTGGAGCACTCCAGGAATGCCATGCAGATT CAGAGTGATAATCTGTACAAATCTGACTTACCAATTTGGATGAAAGGGATCGGCTGGGTGCC ATAGAGTCCCTGGAGGTGGAGAAGGCAAGAAAGCAGGAGAGATTCTTAGTGAGAAGAAGTAT CGCCAGCACCCCGAGAAGCTGAAGTTCACTTACGCCATGGACACAATGGAACAGGCACCTAAC AAGAGTAACAACTGAACATGGACAAGAGGCTCTACACTGAAAAATGGAACAAGGACAAGACC ACCATTATGTCATGCTGACACACCCGATATTTTACTCTCCAGAGTAACCAAAATCACCATG AGTGATAAACTGTACAAAGCTGGCTGGGAAGAGGAAAAGAAGAAAGGATATGACCTGAGGCCT GATGCCATTGCAATAAAGGCTGCAAGAGCTCTAGAGACATTGCCAGTGATTACAAATACAAG AAAGCCTATGAACAAGCCAAAGGGAAACACATTGGCTTCCGGAGCCTGGAAGATGACCCCAAG CTGGTGCACTTCATGCAAGTGGCCAAGATGCAGTCAGACGGGAATACAAGAAGGGATATGAG AAATCCAAGACCTCCTTCCACACCCCGTGGACATGCTCAGTGTGGTGGCAGCCAAGAAGTCT CAGGAAGTGGCCCAATGCCAACTACAGGAACGTGATCCATACCTACAACATGCTTCTTGAT GCCATGAGCTTTGAATTGGCCAAAATATGATGCAGATTCAAAGTGATAATCAGTACAAGGCT GACTATGCTGACTTCATGAAGGGCATTGGATGGCTCCCTCTGGGCTCCCTGGAAGCAGAGAAA AACAAGAAAGCCATGGAGATTATAGTGAAGAAGTACCAGCAAGCAAGCAAGCACTTTGAAG TATTCCCACTCATGGACTCGATGAACATGGTTTTTGGCCAGAAATGCAAAAATATGAAC GAACATCTCTACAAACAGCATGGGAGGCTGACAAAACCAAGTCCACATCATGCCTGATATC CCCCAGATTATTTTGGCAAAGGCAATGCAATTAATATAAGTGATAAACTCTACAACTTTCC
--	--

TTGGAAGAGTCTAAAAAGAAAGGCTATGATCTCAGACCTGATGCAATTCCTATCAAAGCTGCC
AAGGCTTCCAGAGATATTGCAAGTGATTATAAATACAAGTACAATTATGAAAAAGGGAAGGGG
AAAATGGTTGGTTTCCGCAGTCTCGAGGATGATCCCAAATTAGTCCATTCCATGCAAGTGGCT
AAGATGCAATCTGATCGGGAGTACAAGAAAACTATGAGAACAAGACAGCTACACACC
CCTGCCGACATGCTCAGTGTACGGCTGCAAAGGATGCCAAGCCAACATCACAACACTAAC
TACAAGCACCTGATTACAAGTACATCCTCCTCCAGATGCAATGAACATTGAGCTGACCAGG
AATATGAATCGCATACAGAGTGATAATGAATATAAGCAAGATTACAATGAATGGTACAAAGGG
CTTGGCTGGAGTCCAGCAGGTTCTCTGGAAGTGGAGAAGGCCAAGAAAGCAACTGAATATGCC
AGTGATCAGAAATACCGCCAGCACCCGAGCAACTTCCAGTTTAAAGAAGCTGACTGATTCATG
GACATGGTGCTTGCCAAGCAGAATGCACATACCATGAACAAGCATTATACACCATTGATTGG
AATAAAGATAAGACCAAGATTTCATGTGATGCCGTGATACACCAGATATTTTACAAGCCAAGCAG
AATCAAACACTGTATAGTCAGAACTCTATAAACTTGGATGGGAAGAAGCTTTGAAGAAGGC
TATGATCTCCAGTTGATGCAATTTCTGTACAGCTAGCTAAAGCTTCAAGAGACATTGCTAGT
GATTATAAATACAAACAAGGCTACCGAAAGCAACTTGGCCACCATTGTTGGATTCCGGAGTCTG
CAAGATGACCCAAAACCTTGTGTGTCCATGAATGTAGCCAAAATGCAGAGTGAAAGAGAATAC
AAGAAGGACTTTGAGAAGTGGAAAACCTAAGTTCTCCAGCCCAGTGGACATGTTGGGAGTGGTA
CTGGCCAAGAAGTGTGAGGAGTGGTTAGTGACGTGGACTACAAGAATACTGCATCAGTGG
ACATGTCTGCTGTATCAGAACGATGTTGTGCAAGCTAAGAAAGTTTATGAACCTGCAAGTGAG
AATCTATATAAATCTGACCTTGAGTGGCTGAGAGGCATAGGATGGAGTCCCTTGGGTTCTTTA
GAGGCAGAAAAGAACAAAGCGGGCTTCGGAATCATCAGTGAGAAGAAATATCGTCAGCCTCCA
GACAGAAACAAGTTCACCAGCATTCTGATGCCATGGATATAGTTCTGGCAAAGACAAATGCC
AAAAATAGGAGTGATAGACTTTATAGAGAAGCTTGGGACAAGACAAGACTCAGATCCACATC
ATGCCTGATACACCTGACATTGTTCTGGCTAAAGCAAACCTTAATCAACACAAGTGATAAATC
TACCGAATGGGTTATGAGGAGCTGAAGAGAAAAGGTTACGATCTTCTGTTGATGCCATACCA
ATCAAAGCAGCAAAAGCCTCCCGGGAATTGCCAGTGAATACAAGTACAAGGAAGGCTTTCGC
AAGCAGCTCGGCCACCACATTGGTGCCCGAACATTGAAGATGACCCCAAGATGATGTGGTCC
ATGCATGTGGCCAAAGATCCAGAGTGACAGGGAGTACAAGAAGGACTTTGAGAAGTGGAAGACC
AAGTTCAGCAGCCCAGTGGACATGCTGGGGTGGTGTGGCCTATAAGTGCCAGACCTTAGTC
AGCGACGTGGACTACAAGAACTACCTGCACCAAGTGGACATGCTGCGCCAGCAGAGCGATGTC
ATCCATGCTCGGCAGGCCTATGACCTCCAGAGCGATAATTTGTACAAGTCAGACCTTCAGTGG
CTAAAAGGCATTGGCTGGATGACTAGTGGTTCTCTCGAGGATGAGAAAAATAACGAGCCACC
CAGATTTTGAGTGACCATGTTTACCGTCAGCACCCAGATCAATTTAAGTTTTCCAGCCTTATG
GATTCATACCAATGGTTTTGGCAAAAACAATGCTATTACCATGAATCATCGCCTCTATACA
GAAGCTTGGGATAAAGATAAAACCACTGTCCACATTATGCCAGATACCCCTGAAGTTTTATTA
GCTAAACAAAACAAGTAAATTACAGTGAGAAATTGTATAAGCTTGGCCTAGAAGAAGCCAAG
AGGAAAGGTTATGACATGCGGGTAGATGCCATTCCATCAAGGCAGCCAAAGGCTCCAGAGAT
ATTGCAAGTGAATTCAAGTACAAGAAGGCTATCGTAAGCAGCTCGGCCACCACATTGGTGCC
CGAGCTATACGTGATGACCCCAAGATGATGTGGTCCATGCACGTGGCCAAAGATCCAGAGTGAC
AGGGAGTACAAGAAGGACTTTGAGAAGTGGGAAGACCAAGTTCAGCAGCCCAGTGGACATGCTG
GGGGTGGTGTGGCCAAAGAGTGCCAGACCTTAGTCAGCGATGTGGATACAGAAGCTTCCAGT
CACCAGTGGACATGCCTGCCGACCAGAGCGACGTATCATGCTCGGCAGGCCTATGACCTC
CAGAGCGATAATATGTACAAGTCTGATCTCCAGTGGATGAGAGGCATGGCTGGGTGTCCATT
GGCTCTTTGGATGTGGAAAAATGCAAAAGGGCAACTGAAATTTTGAGTGATAAAATCTATCGC
CAGCCTCCAGACAGATTCAAATTTACCAGTGTGACTGACTCTCTGGAACAAGTGTGGCCAAAG
AACAAATGCTCTCAACATGAATAAGCGTTTATACACAGAGGCTGGGACAAAGACAAGACTCAA
ATTACATAATGCCTGATACACCAGAGATTATGTTGGCAAGGCAGAACAAAATCAACTACAGT
GAGACTCTATACAACTTGCCAATGAAGAAGCAAAAAGAAAGGCTACGACTTGCGAAGTGAC
GCCATCCCCATCGTGGCTGCCAAGGCCTCCAGGGACGTTATCAGTGATTACAAATACAAAGAT
GGTTACCGCAAGCAGCTCGGCCACCACATTGGAGCCCGAACATTGAAGATGACCCCAAGATG
ATGTGGTCCATGCATGTGGCCAAGATCCAGAGTGACAGGGAGTATAAGAAGGACTTTGAGAAG
TGGAAGACCAAGTTCAGCAGCCCAGTGGACATGCTGGGAGTGGTGTAGCCAAGAAGTGCCAG
ACCTTAGTCAGCGATGTGGACTACAAGAACTACCTGCACGAGTGGACGTGCCTGCCGACCAG
AATGATGTCATCCATGCTCGGCAGGCCTATGACCTCCAGAGCGATAACATTTACAATCTGAT
CTCCAGTGGCTGAGAGGCATTGGCTGGGTCCCCATTGGGTCTATGGATGTGGTCAAGTGCAAG
AGAGCTGCTGAAATACTGAGTGATAACATCTACCGCCAGCCTCCGGACAAGCTGAAATTTACC
AGTGTGACTGACTCTCTAGAGCAGGTGCTGGCCAAAGAAATGCTCTCAATATGAACAAGCGC
TTATACACAGAAGCCTGGGACAAAGACAAGACCAAGTCCATATTATGCCTGATACACCTGAA
ATCATGTTGGCAAGACAAAATAAAATAAATTATAGTGAGAGCCTCTATCGTCAGGCTGAGAA
GAAGCCAAGAAAGAAGGCTATGACTTGAGAAGTGATGCCATTCCCATTTGTGGCTGCCAAGGCC
TCTCGGGATATTGCCAGTGATTACAAATACAAAGAAGCATATCGTAAGCAGTTGGGTACCAC
ATTGGCGCCCGAGCAGTACACGATGACCCCAAGATAATGTGGTCCCTCCACATTGCCAAAGTG
CAGAGTGACCGTGAGTACAAGAAGATTTTGAGAAATACAAGACAAGGTACAGACGCCAGTG
GACATGCTTGGTATCGTTTGGCCAAAGAGTGTACAGCCTTGGTACAGCATGTGGACTATAAA
CATCCTCTGCATGAATGCATCTGCCTGCCCGACCAGAATGACATCATTCATGCACGGAAAGCC
TATGACCTCCAGAGTGACAATTTGTATAAGTCAAGCCTTGAATGGATGAAAGGCATTGGCTGG
GTTCCGATTGATTCTTGGAAAGTGTAGGGCCAAGAGAGCTGGAGAATTACTTAGTGATACT

ATCTACCGTCAGCGTCCAGAAACGCTGAAATTTACCAGTATAACGGACACTCCGGAGCAGGTG CTGGCAAAAAACAATGCTTTAAACATGAATAAGCGCTTATATACTGAAGCCTGGGACAATGAC AAGAAAATATTTCATGTCATGCCTGATACACCAGAAATCATGTTAGCCAAACTCAACCGAATA AACTACAGTGATAAACTCTATAAACTTGCTTTGGAAGAGTCCAAGAAGGAAGGCTATGACTTG CGTCTGGATGCCATTCCAATCCAAGCAGCCAAGGCTTCAAGAGATATTGCTAGTGATTACAAG TACAAGGAAGGCTACCGCAACAGCTTGGCCACCATTATGGGGCCCGGAACATTAAAGGATGAC CCGAAGATGATGTGCTCCATCCATGTGGCCAAGATCCAGAGTGACAGGGAGTACAAGAAGGAG TTTGAGAAGTGGAAGACCAAGTTTCCAGCAGCCAGTGGACATGCTGGGGGTGGTGCTGGCCAAG AAGTGTCAGATCCTTGTAAGCGACATAGACTACAAGCATCCCTGCATGAATGGACCTGCCTG CCTGATCAGAATGACGTCATTAGGCTCGGAAGGCTATGACCTGCAGAGTGATGCTATTATAC AAATCTGATCTTGAGTGGCTGAGAGGCATAGGATGGGTTCCCATTTGGCTCTGTAGAGGTCGAG AAAGTGAAGAGAGCTGGAGAAATCCTGAGTGACAGGAAGTATCGCCAGCCTGCAGACCAGCTC AAATTCACATGCATTACCGACACTCCGGAATTTGCTTAGCAAAGAATAATGCCCTGACAATG AGCAAGCATTTATACACAGAAGCTTGGGATGCTGACAAAACCTCCATCCAGTGATGCCAGAG ACCCAGATATCCTGCTGGCCAAGAGTAATCTGCCAATATCAGCCAAAACTTTACACCAAG GGATGGGATGAATCAAAGATGAAGGACTATGATCTGAGAGCAGATGCTATTTCCATCAAAGT GCCAAGGCTCCAGGGACATCGCCAGTGACTACAATACAAGGAAGCCTATGAGAAACAGAA GGCCACCACATTGGAGCCAGAGCATTGAAGATGATCCAAGATTATGTGTGCCATACATGCA GAAAAAATTCAAAGTGAAGGGAGTACAAGAAGGAATTCAAAAGTGAAAAACCAAGTTCTCT AGCCCAGTGGACATGTTAAGCATCTTGCTGGCCAAGAAATGTCAGACTTTGGTCACTGACATT TATTATCGCAATTACCTGCATGAATGGACATGCATGCCGGATCAAACGACATTATCCAAGCA AAAAAGGCTATGACCTGCAGAGTGATGCCCTCTACAAGGCTGACTTGGAGTGGTTGCGTGCC ATTGGCTGGATGCCCAAGGGTCTCCTGAAGTGTGAGAGTCAAAAACGCCAGAAATATCTTT TGTGACAGTGTCTATCGGACGCTGTGGTGAACCTTAAGTACACAAGCATTGTTGACACACCT GAAGTGGTCTCTGCTAAATCAAATGCTGAAAATATTAGTATTCCAAAGTACAGAGAGGTTGG GACAAGGATAAACTTCAATACACATAATGCCAGATACTCCAGAAATTAATCTCGTAGAGCA AATGCTCTTAATGTGAGCAATAAACTTTACCGTGAGGGCTGGGATGAAATGAAGGCGTCTGT GATGTCCGGCTGGATGCCATCCCATCCAGGCTGCCAAGGCTCCAGGGAGATTGCCAGTGAC TATAAATATAAGCTTGACCATGAGAAGCAGAAGGGACACTACGTGGGCACCTCAGAGCCAGG GATGACAACAAGATCCGCTGGGCCCTCATAGCTGACAAGCTCCAGAAATGAACGAGAGTACCGG CTGGACTGGGCCAAATGGAAGGCCAAGATCCAGAGCCCTGTGGACATGCTTTCCATCCTGCAC TCTAAAAATTCAGGCTCTGGTTCAGTGACATGGATTACCGCAATTACCTGCACAGTGGACC TGCATGCCCGACCAGAACGATGTGATTCAGGCCAAGAAGGCTACGAACTGCAGAGCGATAAT GTTTACAAGGCTGACTTGAATGGTTGCGTGGAATGGGTGGATGCCAAATGACTCCGTGTCC GTCAATCATGCCAAACATGCCCGGACATCTTCAGTGAGAAAAATATCGCACAAAAATAGAA ACTCTCAACTTTACGCCCTGTGGATGACAGAGTTGATTATGTGACAGCGAAACAAAGTGGCGAG ATCCTCGATGATATTAATACCGGAAAGACTGGAATGCCACCAATCAAAGTACACCCTCACA GAAACCCCTGCTGCACACTGCCAGGAGGCTGCTAGGATACTGGACCAGTATCTCTACAAG GAAGGCTGGGAGAGACAAAAGCCACAGGTTACATTTTGCTCCAGATGCTGTGCCATTGTT CATGCCCATCACTGCAATGACGTTTACAGTGAGCTGAAATACAAGCTGAACATGTAAGCAA AAAGGTCATTATGTTGGTGTCCGACGATGAGAGATGATCCTAAGCTGGTTGGTTGAGCAT GCAGGCCAGATTGAGAAATGAGAGACTATACAAGAGGACTATCACAACAAAGGCCAAATC AATATACCTGCTGATATGGTGTGCTGCTTGGCCGCAAGCAGGGGAGACCTTGTGAGTGAT ATTGATTATCGTAATTACTTGCACCAATGGATGTGTCATCCTGACCAGAACGATGTTATTACG GCAAGAAAGGCTATGACCTACAGAGTGATAATGTCTACAGAGCTGACCTGGAGTGGCTCCGA GGCATTGGCTGGATCCCACTGGATTCTGTGGACCATGTAAGGGTTACTAAGAACCAGGAAATG ATGAGTCAGATCAAATATAAGAAAAATGCCCTTGAAAACATATCCTAATTTACAAGTGTGGTG GATCCTCCAGAGATTGTTTTAGCCAAGATTAATCTGTCAATCAAAGTGATGTAATAATATAAA GAAACATTTAATAAAGCAAAGGCCAAATATAACGTTTTACCAGATACACCACATATCTCCAC TCCAAGACATGGGAAAACTCTACAGTACTATACTGTATAAAGGGGCTGGGAGGGCACCAG GCCTATGGCTACACCCTGGATGAGCGCTACATTCCTTGTGGAGCCAAGCATGCTGATCTG GTGAACAGTGAGCTTAAATACAAGAGACATATGAGAAGCAGAAGGTCATACCTGGCTGGA AAAGTGATCGTGAAATCCCTGGTGTGGTTCACTGTCTGGATTTCAAAAGATGAGGAGTGCG TTGAACTACAGAAAAATTATGAGGATACCAAGCAAATGTTTCATATCCCAATGACATGATG AATCACGTGCTGGCTAAAAGGTGCCAGTACATCCTCAGTGACCTGGAGTATCGACACTATTTC CACCAGTGGAGCTCTCTTCTGGAAGAACCAATGTTATACGCGTCCGAAACGCCAGGAGATC TTGAGTGATAATGTGTATAAAGATGACCTGAATGGTTGAAAGGCATTGGTTGCTACGTTTGG GATACACCCCAATCCTCCATGCCAAGAAATCATACGACCTTCAGAGTCAGCTACAATATACA GCAGCAGGTAAGAAAAATCTACAAAATATAATCTGGTTCACAGACACGCCCTCTATGTGACT GCTGTTTACAGTGGCATTAAATGCCAGTGAGGTAATAATAAGAAAAATATCATCAGATTAAAG GACAAATACACAACAGTTCTAGAAACAGTGGATTATGACAGAACCAGAAACCTGAAGAATCTT TACAGCAGTAACCTGTACAAGGAGGCTGGGATAGAGTGAAAGGCCACAGCTACATCTGCCCT TCCAGCCTTGTCTCTGACACACGCCAAGAACCAAGCATCTGGCCAGCCATATCAAATAT CGGGAAGAATATGAAAAGTTCAAAGCTCTTTATACGTTACCAAGAAGTGTGACGATGATCCG AACACAGCAGCGTGCCTCCGAGTTGGCAAGCTTAACATCGATCGCTGTACAGATCAGTTTAT GAAAAGAACAAGATGAAAATCCACATCGTGGCCGACATGGTAGAGATGGTTACTGCCAAGGAT
--

TCCCAGAAGAAAGTCAGTGAGATTGATTACCGCCTGCGCCTCCACGAATGGATTGGCCACCCC
GACTTGCAAGTCAATGATCACGTGAGGAAAGTCACAGATCAGATCAGCGATATTGTATACAAG
GATGACCTCAACTGGCTGAAAGGCATTGGTTGCTACGTCTGGGACACTCCTGAAATCCTCCAT
GCCAAGCATGCTTATGATCTACGTGATGATATCAAGTATAAAGCTCACATGTTGAAACAAGG
AATGACTACAAGCTTGTCACAGATACACAGTCTACGTGACAGCTGTCAAAAGTGGGAAACAG
CTAAGTGACGCTGTCTACCACTATGACTATGTGCACAGTGTGAGAGGCAAAGTGGCTCCAACT
ACCAAAACCGTGGATCTGGACCGGGCCCTTCATGCATACAAGCTCCAGAGTTCGAATCTATAC
AAAACAGCCTGCGCACCCCTGCCCCTGGATATAGACTTCCAGGTGACACTCCTCACTTCAAA
CACATCAAGGACACCCGTTACATGAGCAGTTATTTCAAGTACAAAGAAGCCTATGAACACACC
AAGGCATATGGGTATACACTTGGCCCCAAAGATGTTCCATTTGTCCACGTCCGGAGAGTCAAC
AATGTTACAGCGAGAGACTGTATCGGGAATGTACCAAACTGAAAGACAAGATCCATACA
ACTCCCGATCCCCCTGAGATCCGCCAAGTCAAGAAGACACAAGAGGCTGTGAGTGAGTTGATC
TACAAATCAGACTTCTTCAAGATGACGGCCACATGATCTCTGCCATACACACCCCAAGTG
ATCCATTGCCGCTATGTGGGAGACATCACCAGTGATATTAAATACAAAGAGGACTTGCAGGTC
CTGAAGGGATTGGCTGCTTCTGTATGACACTCCTGACATGGTCCGCTCCCGGCACCTGCGG
AAGCTCTGGTCTAATTACCTATACACTGATAAGGCAAGGAGATGCGAGACAAATACAAAGTG
GTGCTTGACACTCCAGAATACAGAAAAGTGAAGAAGTGAAGACACATCTGAGTGAGCTGGTC
TACAGAGCTGACAGCAAGAAGCAGAAGTCAATCTTTACTTCAGTTCTGTGATCTCTGATCTT
TTAAGAGCCAAGCGAGGGCAGAAGCTTCAGAGTCAGTATCTGTATGTTGAACTTGCCACCAA
GAGAGACCCCATCATCACGTGGAACACAGACCACAGCCTTGAAGCATGCTAAAGACGTGAAG
GACATGGTCAAGTGAGAAAAGTACAAGATTCAATATGAAAAGATGAAAGACAAGTACACTCCG
GTCCAGATACGCCAATCCTCATCAGAGCCAAGAGGCTTACTGGAATGCCAGTGATCTACGC
TACAAAGAAACATTTCAAAGACCAAAGGAAATACACACGGTGAAGATGCCCTAGACATT
GTCTATCATCGCAAAGTCACAGATGACATCAGTAAATAAAATACAAAGGAGAACTACATGAGC
CAGTTGGGTATCTGGAGGTCCATTCTGTATCGTCCAGAGCATTTCCACCACCGAGCAGTCACT
GACACAGTCAGTGATGTAAATATAAAGAAGACTTGACTTGGCTTAAAGGCATTGGTTGCTAT
GCCATGATACCCCTGATTCTACTCTGGCTGAAAAGAACAAAGACTCTACAGCAAGTATAAG
TATAAAGAAGTATTTGAAAGGACAAAGTCAGATTTCAAGTATGTTGCCGACTCTCCGATCAAT
AGGCATTCAAGTATGCAACTCAATTGATGAATGAGAAAAATACAGAGCTGATTATGAGCAG
CGGAAAGATAAATACCACCTGGTAGTCGATGAGCCTAGACATCTGCTGGCTAAGAACCCGACG
GACCAGATCAGTCAGATCAATACAGGAAAACTATGAAAAATCAAAGGACAAATTTACCTCA
ATTGTGGATACTCCAGAACACCTGCGTACTACAAAGTCAACAAACAAATCAGCGATATCCTT
TATAAATTGGAATACAACAAGGCCAAACCCAGAGGCTACACCACAATCCACGACACGCCCATG
TTGCTGCATGTCCGAAGGTTAAGATGAAGTCAGTGATCTGAAATACAAAGAAGTATACCAA
AGAAATAAATCCAATGCACCATTGAGCCAGATGCTGTTCAATATCAAAGCAGCCAAAGGACGCC
TACAAAGTCAACACCAATCTGGACTATAAGAAACAGTACGAAGCCAAACAAAGCCACTGGAAG
TGGACTCCTGACCGACCGGACTTCTCCAGGCTGCCAAGTCATCCTGCAGCAAAGCGATTTT
GAATATAAGCTGGACCGGAGTTCTCAAGGGTTGCAAGCTTTCTGTCACTGATGACAAAAAC
ACGGTGCTCGCCCTCAGGAATACTTTAATAGAAAGTATCTGAAATACAAAGAGAAACATGTC
AAGGAAAGAGGAACCTGTCTGCGGTACCTGACACGCTCAGATCCTGTGGCGAAGCTGTC
AGCAACCTGGTGTCTGAGAACAAAGTACAAGGACCATGTCAAGAAGCACTTGGCACAGGGCTCA
TACACAACACTACCAGAGACCCGGGACACTGTTACGTCAAGGAAGTGACCAAGCATGTCACT
GATACAAATTACAAAAAGAGTTTGTCAAGGAGAAAGGAAAAATCCAATCTCCTATGCTG
GAGCCACAGAGGTGAAACATGCTATGGAAGTGGCCAAGAAGCAAAGTGATGTCGCTTACAGA
AAAGATGCCAAAGAGAAAGCTGCATTACACACAGTGGCTGATCGACCAAGATCAAGAAGGCC
ACACAGGCAGCCAAACAGGCCAGTGAGGTGGAGTACAGAGCCAAGCACCGCAAGGAAGGCAGC
CATGGCTTAAGCATGCTCGGTCGCCAGACATAGAAATGGCCAAGAAGGCAGCCAAAGCTGAGC
AGCCAGGTTAATACCGAGAAAAATTCGATAAAGAAAAGGGCAAGACACCAAAATACAATCCA
AAAGACAGCCAGCTCTACAAAGTCATGAAAGATGCTAATAATCTTGCAAGTGAGGTTAAATAC
AAGGCTGACCTGAAGAACTTCAAAACCCGTGACTGACATGAAGGAGTCTCTGATCATGAAT
CATGTCTGAATACAAGCCAATCTGCCAGTTCTTACCAGTACAAGAAGAGTATGAGAAGAGT
AAAGGCCACTACCACACCATAACCGATAATCTGGAGCAGCTTCACTTAAAGAGGCCACAGAA
TTACAGAGTATAGTGAATACAAAGAAAAGTATGAAAAGGAACGAGGAAAAACCATGTGGAC
TTTGAACACCAACGTACATCACTGCCAAAGAGTCTCAGCAGATGACAGTGGGAAAGAAATAT
AGGAAAGATTATGAAGAGTCCATTAAAGGCAGAAACCTGACTGGCCTGGAGGTACGCCAGCT
TTGTTACATGTCAAATATGCAACTAAATAGCAAGCGAGAAAGAGTACAGGAAAGATCTAGAG
GAAAGCATCCGTGGGAAGGGCCTCACTGAAATGGAAGATACACCTGACATGCTAAGAGCAAAG
AATGCCACTCAAATCCTCAATGAGAAAGAAATATAAGCGAGACCTGGAAGTCAAGGA
AGAGGCCTGAATGCCATGGCCAAATGAACTCCGGATTTTATGAGGGCCAGGAATGCTACTGAT
ATTGCCAGTCAGATTAAGTATAAGCAATCAGCAGAAATGGAGAAAGCCAAATTTCACTTCTGTG
GTTGATACTCCAGAGATCATTATGCCCAACAAAGTCAAGAATCTTTCAAGCCAGAAAAAGTAC
AAGGAAGATGCTGAGAAGTCCATGTCGTATTATGAGACTGTTTGGACACCCAGAGATACAG
AGAGTCCGGGAGAACCAAGAACTTCAAGCTTCTCCAATACCAGTGTGACCTTAAAGACAGT
AAAGGAAAAATTACAGTTGTTCAAGACACGCCAGAAATCTGCGTGTAAGAAAGAAATCAGAAG
AATTTCAAGCTCGGTTTATATAAAGAGGATGTCTCACCAGGAACGGCTATCGGAAGACACCT
GAGATGATGAGAGTGAACCAACACAGGACCACATTAGCTCGGTGAAGTATAAGGAAGCAATA

	<p>GGACAAGGAAGTCCAATCCCTGACCTGCTGAAGTGAACGTTGTAAGGAGACGCAGAAGCAC ATTAGCTCGGTTATGTACAAAGAAAAGTGGGAACAGGCATTCCAACCACTGTGACTCCAGAG ATTGAGAGAGTCAAACGCAATCAAGAGAACTTTAGCTCGGTTTTGTACAAAGAAAATGTGGGG AAAGCCACCGCAACCCCTGTCACTCCTGAGATGCAGAGAGTCAAACGCAATCAAGAAAACATT AGCTCGGTATTGTACAAAGAGAACATGAGAAAAGCAACTCCGACACCTGTTACTCCAGAGATG GAGAGAGCTAAGCGCAACCAAGAAAACATTAGCTCGGTTCTTTATTCTGATAGTTTCCGAAAA CAAATACAAGGCAAAGCTGCCTATGTATTGGATACCCCCGAGATGAGACGGGTGAGGGAGACC CAACGGCACATCTCAACGGTGAAATATCATGAAGACTTTGAGAAACACAAGGGTGTCTTCACA CCAGTGGTGACAGATCCTATCACTGAACGAGTAAAGAAGAACATGCAGGACTTCAGTGACATT AACTACCGAGGTATTGAGAGGAAAGTGGTAGAAATGGAACAAAAACGGAATGACCAAGATCAG GAGACTATTACAGGTTTACGTGTCTGGCGTACTAATCCTGGTTCGGTTTTGACTATGATCCA GCAGAAGACAACATCCAGTCCCGAAGCTTACACATGATTAATGTCCAAGCTCAGCGCCGGAGC CGGGAGCAGTCACGATCTGCGAGTCACTAAGCGTCAGTGGGGTGAGGAGAAGTCTGAGCAT TCAGAAGCACCAGACCACCACCTTTCGACTTACAGCGACGGGGGTGCTTTGCGATCTCAACA GCTTACAAACATGCAAAAACCACAGAGCTCCCAACAACGATCATCTTCAGTTGCTACCCAA CAGACAACGGTATCTTCCATCCCATCTCATCCATCTACTGCTGGAATAATCTTCCGTGCCATG TATGACTATATGGCTGCTGATGCAGATGAGGTGTCCTTCAAGGATGGAGATGCCATCATAAAT GTCAAGCAATTGATGAAGGCTGGATGTATGGCACTGTGCAGAGGACTGGCAGGACCGGAATG CTCCAGCCAACCTACGTTGAAGCTATTTAGGCATTTCAAAGCATCACACTTGTCTGCAGGACT TACAGATCTGCGAGTCAATGTTTCGGTTTAGACTCTCCACTGTTACCTAAGTTCTCAAGCTGC CTATGGTTTTTCTGTGCTCAATGTGATTTATGGTAGTACCATCCTTTCTCCTTGGGTTTTAAA ATAAGTTGCAGAACAGACACTTTAAAGCTTCTGCAATATTATTTCTGTGCTAGAGTCTTTT TCCATTATAAACATGTTTTTAACATTATTTCTTTTCTAAACAGGGATTTTGAATATGCCAAAC ACATTAAAGGAAAAATAGCAGAGATGTTTCACTTTTCTTCTGCTGATTGCTAATGCTTATTATT TCTAATTGAGTTCTGAAGTTATAAAGTTATAATCAATACAAACCAGCAACTAATAAAACCTCT AATTCTGCAAAAAAAAAAAAAAAAAAAGTCG</p>
	<p>ORF Start: ATG at 19</p>
	<p>ORF Stop: TAG at 19747</p>
	<p>SEQ ID NO: 144</p>
	<p>6576 aa</p>
	<p>MW at kD</p>
<p>NOV35c, CGI19566-03 Protein Sequence</p>	<p>MADDEDYEEVVEYYTEEVVEYVGETITKIYETTTTTSYEQSETSKPALAQPALAQPASA KPEVERRKVIKKVDPKFMTPYIAHSQKMDLFSFNKYKEFEKTKGQPYASTTDTPELRRIK KVQDQLSEVKYRMDGDVAKTICHVDEKAKDIEHAKKVSQVSKVLYKQNWEDTKDKYLLPPDA PELVQAVKNTAMFSKKLYTEDWEADKSLFYPNDSPELRRVAQAQKALSDVAYKKGLAEQAQ FTPLADPPDIEFAKKVTNQVSKQYKEDYENKIKGKWSETPCFEVANARMNADNISTRKYQED FENMKDQIYFMQTETPEYKMNKAGVAASKVKYKEDYENKKGADYNVLPAENPQLRQLKAA GDALSDKLYKENYEKTKAKSINCYETPKFKLDTVLQNFSSDKKYKDSYKLDILGHYVGSFEDP YHSHCMKVTAQNSDKNYKAEYEEDRGKGFPTITQEYDAIKKLDQCKDHTYKVPDKTKFTQ VTDSPVLLQAQVNSKQLSDLNKAKHESEKFKCHIPPDTPAFIQHKVNAYNLSNLKQDWEK SKAKKFDIKVDAIPLLAANKANTNTSDVMYKKDYKNGKMGVLSINDDPKMLHSLKVAKNQ SDRLYKENYEKTKAKSMNYCETPKYQLDTQLKNFSEARYKDLVYKDVILGHYVGSMEPDYHTHC MKVAAQNSDKSYKAEYEEDKGKCYFPQITQEYDAIKKLDQCKDHTYKVPDKTKFTAVTDS VLLQAQVNSKQLSDLNKAKHEGERFKCHIPADAPQFIQHRVNAYNLSNLKQDWEKSKAKK FDIKVDAIPLLAANKANTNTSDVMYKKDYKSKGKMIGALSINDDPKMLHSLKTAQNQSDREY RKDYKSKTIYAPLDMQVTAQKSKQAIASDVYKHILHSYSYPPDSINVDLAKKAYALQSD VEYKADYNSWMKGCWVPPFSGLEMEKAKRASDILNEKKYRQHPDTLKFSTIEDAPITVQSKIN QAQRSDIAYKAGGEEIHNLYNLPDLPQFIQAKVNAYNISENMYKADLKDLSKKGYDLRTDAI PIRAAKAARQAASDVQYKDYKAKGKMVGFSQDQDPKLVHYMNVAKIQSDREYKDYKTK SKYNTPHDMFNVVAAKQADVVSNVNYKHSLLHYTYLPDAMDLELSKNMMQIQSDNVYKEDYN NWMKGIGWIPIGSLDVEKVKAGDALNEKKYRQHPDTLKFSTIVDSPVMVQAKQNTKQVSDIL YKAGGEDVKHKYTMSPDLQFLQAKCNAYSISDVYKRDWDLIRKGNVVGDAIPITAAKAS RNIAADYKYKEAYEKSKGKHVGFSLQDDPKLVHYMNVAKIQSDREYKKNYENTKTSYHTPGD MVTITAAKMAQDVATNVNYKQPLHHTYTYLPDAMSLEHTRNVNQIQSDNVYKDEYNSFLKGIGW IPIGSLEVEKVKAGDALNERKYRQHPDTVKFTSVDPDSMGMLAQHNTKQLSDLNKVEGEKL KHKTYIDPELPQFIQAKVNALNMSDAHAKDWKKTIRKGYDLRPAIPIVAAKSSRNIAADCK YKAEYKAKGKQVGFSLQDDPKLVHYMNVAKIQSDREYKGYEASKTYHTPLDMVSVTAAK KSQEVATNANYRQSYHHYTLPLDALNVEHSRNAMQIQSDNLKSDFTNWMKGIGWVPIESLEV EKAKKAGEILSEKKYRQHPKLFYAMDTMEQALNKNLMDKRLYTEKWNKDKTTIHVMP DTPDILLSRVNIQIMSKLYKAGWEEKKKGVDLRPAIAIKAAASRDIAADYKYKAYEQA KKGHIGFRSLEDDPKLVHFMQVAKMQSDREYKKGYSKTSFHTPVDMLSVVAAKKSQEVATN ANYRNVITHYTNLPDAMSFLAKNMMQIQSDNQYKADYADFMKGIGWLPGLGSLEAEKNKKAME IISEKKYRQHPDTLKYSTLMDSMNMLAQNNAKIMNEHLYKQAWADTKVHIMPDIPIQIL KANAINISDKLYKLSLEESKKKGVDLRPAIPIKAAASRDIAADYKYKAYEQA KYLILLPDAMNIELTRNNRIQSDNEYKQDYNEWYKGLGWSAGSLEVEKAKKATEYADQKYR QHPSNFQFKLTDMSDMVLAKQNAHTMKNHLYTIDWNKDKTIHVMPDTPDILQAKQNTLYS QKLYKLGWEEALKKGYDLPVDAISVQLAKASRDIAADYKYKQYRQKLGHHVGFSLQDDPKL</p>

VLSMNVAKMQSERYKKDFEKWTKFSSPVDMLGVVLAKKCQELVSDVDYKNYLHQWTCPLDQ
 NDVVQAKKVYELQSENLYKSDLEWLRGIGWSPLGSLEAEKNKRASEI ISEKKYRQPPDRNKFT
 SIPDAMDIVLAKTNAKNRSRLYREAWDKDKTQIHIMPDPDIPVLAKANLINTSDKLYRMGYE
 ELKRKGVDLPVDAIPIKAASREIASEYKKEGFRKQLGHHIGARNIEDDPKMMWSMHVAKI
 QSDREYKKDFEKWTKFSSPVDMLGVVLAYKCQTLVSDVDYKNYLHQWTCPLDQSDVIHARQA
 YDLQSDNLYKSDQLWLGIGWMTSGSLEDEKNKRATQILSDHVYRQHPDQKFSSLMDSIPMV
 LAKNNAITMNHRLYTEAWDKDKTTHIMPDPTEVLLAKQNKVNYSEKLYKLGEEAKRKGDM
 RVDIAPIKAASRDIASEFYKKEGYRKQLGHHIGARAI RDPKMMWSMHVAKI QSDREYKKD
 FEKWKTKFSSPVDMLGVVLAKKCQTLVSDVDYKNYLHQWTCPLDQSDVIHARQAYDLQSDNMY
 KSDLQWMRGIGWVSIGSLDVEKCKRATEILSDKIYRQPPDRFKFTSVTDSLEQVLAKNNALNM
 NKRLYTEAWDKDKTQIHIMPDPTEIMLARQNKINYSLEYLANEEAKKGYDLRSDAIPIVA
 AKASRDVSDYKYKDGFRKQLGHHIGARNIEDDPKMMWSMHVAKI QSDREYKKDFEKWTKFSS
 SPVDMLGVVLAKKCQTLVSDVDYKNYLHEWTCPLDQNDVIHARQAYDLQSDNLYKSDQLWLRG
 IGWVPIGSMDDVVKRAAEILSDNIYRQPPDKLFTSVTDSLEQVLAKNNALNMNKRLYTEAW
 DKDKTQVHIMPDPTEIMLARQNKINYSLEYRQAMEEAKKEGYDLRSDAIPIVAASRDIA
 DYKYKEAYRKQLGHHIGARAVHDDPKIMWSLHIAKVQSDREYKKDFEKYKTRYS SPVDMLGIV
 LAKKCQTLVSDVDYKHPHECICLPDQNDI HARKAYDLQSDNLYKSDLEWMLGIGWVPIDSL
 EVVRAKRAGELLSDTIYRQRPETLFTSITDTPQVLAKNNALNMNKRLYTEAWDKDKTTHV
 MPDTPPEIMLAKLRNINYSKLYKLALESKEGYDLRLDAIPIQAASRDIA SDYKYKEGYR
 KQLGHHIGARNIKDDPKMMWSIHVAKI QSDREYKKEFEKWKTKFSSPVDMLGVVLAKKCQILV
 SDIDYKHPHEWTCPLDQNDVIQARKAYDLQSDAIYKSDLEWLRGIGWVPIGSVEVENVKRAG
 EILSDRKYRQPADQLKFTCITDTPPEIVLAKNNALTM SKHLYTEAWDADKTSIHVMPDTPDILL
 AKSNSANISQKLYTKGWDESKMKDYDLRADAISI KSAKASRDIA SDYKYKEAYEKQKHHIGA
 QSIEDDPKIMCAIHAEKI QSEREYKKEFQKWKTKFSSPVDMLSILLAKKCQTLVTDIYRNYL
 HEWTCMPDQNDIIQAKKAYDLQSDALYKADLEWLRGIGWMPQGSPEVLRVKNQNFCDSDVYR
 TPVVNLKYTSIVDTPPEVLAKSNAENISIPKYREVWDKDKTSIHIMPDPTEINLARANALNV
 NKLYREGWDEMKGACDVRDAIPIQAASREIASDYKYKLDHEKQKGHYVGTLTARDNDKIR
 WALIADKLQNEREYRLDWAQWAKI QSPVDMLSILHSKNSQALVSDMDYRNYLHQWTCMPDQ
 DVIQAKKAYELQSDNLYKADLEWLRGIGWMPNDSVSVNHAKHAADIFSEKKYRTKIETLNF
 VDDRVDYVTAKQSGEILDDIKYRKDNATKSKYTLTETPLLHTAQEAARILDQYLYKEGWERO
 KATGYILPPDAVPFVHAHCHNDVQSELKYKAHVQKQGHYVGVPTMRDPLKLEWDFHAKQIQN
 ERLYKEDYHKTAKINIPADMVSVLAAKQGQTLVSDIDYRNYLHQWMCMPDQNDVIQARKAYD
 LQSDNLYRADLEWLRGIGWIPLDSDVHVRVTKNQEMMSQIKYKKNALENYPNFTSVDPPEIV
 LAKINSVQSDVYKETFNKAQKGYTFSPDTPHISHSKDMGKLYSTILYKGAWEKTKAYGYTL
 DERYIPIVGAHADLVNSELKYKETEYKQKGHYLAGKVIGEFPGVVHCLDFQKMRSAALYRKH
 YEDTKANVHIPNDMMNHVLAQRQYILSDLEYRHYFHQWTSLLPEPNVIRVRNAQEILSDNVY
 KDDLNLWLGIGCYVWDTPOILHAKKSYDLQSQLQYTAAGKENLQNYNLVTDTPLYVTAVQSGI
 NASEVKYKENYHQIKDKYTTVLETVDYDRTRNLKNLYSSNLYKEAWDRVKATSYILPSTLSL
 THAKNQKHLASHIKYREEYEFKALYTLPRSVDDDPNTARCLRVGKLNIDRLYRSVYEKNKM
 IHIVPDMVEMVTAKDSQKKVSEIDYRLRLHEWICHDPDLQVNDHVRKVTDQISDIVYKDDLNLW
 KGIGCYVWDTPEILHAKHAYDLRDDIKYKAHMLKTRNDYKLVTDTPVYVQAVKSGKQLSDAVY
 HYDYVHSVRGKVAPTTKTVDLRALHAYKLQSSNLYKTSRLTLPTGYRLPGDTPHFKHIDTR
 YMSSYFYKKEAYEHTKAYGTYLGPDKDVPFVHVRVNNVTSERLYRELYHKLKDKIHTTPDPPE
 IRQVKKTQEA VSELIYKSDFFKMQGHMISLPYTPQVIHCRYVGDITS DIKYKEDLQVLKGF
 FLYDTPDMVRSRHLKLSNLYTDKAREMRDXYKVVDLTPYRKYQELKTHLSLYRAAGK
 KQKSIFTSVPDTPDLLRAKRGQKLQSQYLYVELATKERPHHHAGNQTTALKHAKDKVMVSEK
 KYKIQYKMKDKYTPVPDTPILIRAKRAYWNASDLRYKETFPQKTKGYHTVKDALDIVYHRKV
 TDDISKIKYKENYMSQLGIWRSIPDRPEHFHRAVTDTVSDVYKEDLTWLKGIGCYAYDTPD
 FTLAEKNKTLYSKYKKEVFERTKSDFKYVADSPINRHFYATQLMNEKKYRADYEQRKDKYH
 LVVDEPRHLLAKTRSDQISQIKYRKNYEKSKDKFTSIVDTPPEHLRTTKVNQKISDILYKLEYN
 KAKPRGYTTIHDTMPLLHVRKVDEVSDLYKKEVYQRNKSNTIEPDVHIAKADAYKVNTN
 LDYKKQYEAANKAHWKWTPDRPDFLQAAKSSSQSDFEYKLDREFLKGCKLSVTDDKNTVLALR
 NTLIESDLKYKEKHVKERTCHAVDTPQILLAKTVSNLSENKYKDHVKKHLAQGSYTTLPE
 TRDTHVHKEVTKHVSDTYKKFKGKSNYSIMLEPPEVKHAMEVAKKQSDVAYRKDAKEK
 LHYYTVADRDPDIKKATQAAKQASEVEYRAKHRKEGSHGLSMLGRPDIEMAKKAALSSQVKYR
 ENFDKEGKTPKYNPKDSQYLYKVMKDANNLASEVKYKADLKKLHKPVTDKESLIMNHVNLTS
 QLASSYQYKKYKESKGYHTIPDNLEQLHLKEATELQSVYKKEKYEKERGKPMLEFETPT
 ITAKESQMQSGKEYRKDYEEISIGRNLTGLEVTALLHVKYATKIASEKEYRKOLEESIRGK
 GLTEMEDTPMLRAKNATQILNEKEYKRDLELEVKGKRLNANANETPDMFRARNATDIA SQIK
 YKQSAEMEKANFTSVVDTPEIIHAQVKNLSSQKYEKEDAEKSMSYETVLDTPEIQRVRENO
 KNFSLQYQCDLKNKSGKITVQDTPPEILRVKENQKNFSSVLYKEDVSPGTAIGKTPEMMRVK
 QTQDHISVYKKEAIGQGTPIPDLPVVKRVKETQKHISVYKENLGTGIPTPVTEIERVKR
 NQENFSSVLYKENVKGATATPVTPEMQRVKRNQENISSVLYKENMRKATPTPTVTEIERAKRN
 QENISSVLYSDSFRKQIQGAAYVLDTPEMRRVRETQRHISTVKYHEDFEKHKGCFTPVVTD
 ITERVKKNMQDFSDINRYGIQRKVVEQKRNQDQETITGLRVWRNTPGVSFVDPYPAEDNIQ
 SRSLHMINVQAQRRSREQRSASALSVSGGEEKSEHSEAPDHLSTYSDGGVFAVSTAYKHAK

TTELPQQRSSSVATQQTTVSSIPSHPSSTAGKIFRAMYDYMAADAEVSFKDGDALINVQAIDE GWMYGTVQRTGRTGMLPANYVEAI

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 35B.

Table 35B. Comparison of NOV35a against NOV35b and NOV35c.		
Protein Sequence	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV35b	24..6700	6453/6677 (96%)
	24..6669	6456/6677 (96%)
NOV35c	24..6400	6165/6386 (96%)
	24..6405	6185/6386 (96%)

5

Further analysis of the NOV35a protein yielded the following properties shown in Table 35C.

Table 35C. Protein Sequence Properties NOV35a	
PSort analysis:	0.8800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV35a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 35D.

15

Table 35D. Geneseq Results for NOV35a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG01254	Novel human diagnostic protein #1245 - Homo sapiens, 1573 aa. [WO200175067-A2, 11-OCT-2001]	740..1871 405..1518	482/1161 (41%) 756/1161 (64%)	0.0
ABB12450				e-136

	protein SEQ ID NO: 289 - Homo sapiens. 1654 aa. [WO200174836-A1, 11-OCT-2001]	1..432	336/500 (67%)	
AAM40050	Human polypeptide SEQ ID NO 3195 - Homo sapiens, 370 aa. [WO200153312-A1, 26-JUL-2001]	3333..3660 1..328	148/331 (44%) 226/331 (67%)	4e-79
AAM28003	Peptide #2040 encoded by probe for measuring placental gene expression - Homo sapiens, 102 aa. [WO200157272-A2, 09-AUG-2001]	4008..4109 1..102	102/102 (100%) 102/102 (100%)	5e-55
AAM15512	Peptide #1946 encoded by probe for measuring cervical gene expression - Homo sapiens, 102 aa. [WO200157278-A2, 09-AUG-2001]	4008..4109 1..102	102/102 (100%) 102/102 (100%)	5e-55

In a BLAST search of public sequence databases, the NOV35a protein was found to have homology to the proteins shown in the BLASTP data in Table 35E.

Table 35E. Public BLASTP Results for NOV35a

Protein Accession Number	Protein/Organism/Length	NOV35a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P20929	Nebulin - Homo sapiens (Human), 6669 aa.	1..6700 1..6669	6603/6705 (98%) 6619/6705 (98%)	0.0
Q14215	NEBULIN - Homo sapiens (Human), 3007 aa (fragment).	3573..6561 4..2980	2668/3012 (88%) 2770/3012 (91%)	0.0
Q14214	NEBULIN - Homo sapiens (Human), 2472 aa (fragment).	280..2750 1..2471	2460/2471 (99%) 2466/2471 (99%)	0.0
Q9DEH4	Nebulin - Gallus gallus (Chicken), 2402 aa (fragment).	4011..6381 1..2339	1572/2389 (65%) 1878/2389 (77%)	0.0
Q62411	Nebulin - Mus musculus (Mouse), 1358 aa (fragment).	2116..3473 1..1358	1238/1358 (91%) 1311/1358 (96%)	0.0

5

PFam analysis predicts that the NOV35a protein contains the domains shown in the Table 35F.

Table 35F. Domain Analysis of NOV35a

Pfam Domain	NOV35a Match Region	Identities/ Similarities for the Matched Region	Expect Value
-------------	---------------------	---	--------------

Nebulin_repeat	82..110	14/29 (48%) 24/29 (83%)	0.0011
Nebulin_repeat	118..146	12/29 (41%) 25/29 (86%)	1.1e-07
Nebulin_repeat	153..181	10/29 (34%) 25/29 (86%)	8e-05
Nebulin_repeat	188..216	13/29 (45%) 24/29 (83%)	5e-08
Nebulin_repeat	223..251	9/29 (31%) 24/29 (83%)	0.0022
Nebulin_repeat	258..286	14/29 (48%) 25/29 (86%)	4.9e-07
Nebulin_repeat	293..321	11/29 (38%) 21/29 (72%)	0.00095
Nebulin_repeat	329..357	17/29 (59%) 24/29 (83%)	8.3e-08
Nebulin_repeat	368..396	11/29 (38%) 26/29 (90%)	2.7e-07
Nebulin_repeat	439..467	9/29 (31%) 23/29 (79%)	0.013
Nebulin_repeat	507..535	12/29 (41%) 25/29 (86%)	3e-06
Nebulin_repeat	542..570	14/29 (48%) 26/29 (90%)	2.1e-10
Nebulin_repeat	578..606	16/29 (55%) 26/29 (90%)	4.7e-09
Nebulin_repeat	616..644	15/29 (52%) 28/29 (97%)	3.6e-10
Nebulin_repeat	686..714	11/29 (38%) 22/29 (76%)	0.0072
Nebulin_repeat	754..782	11/29 (38%) 25/29 (86%)	8.2e-05
Nebulin_repeat	789..817	12/29 (41%) 26/29 (90%)	4e-09
Nebulin_repeat	825..853	16/29 (55%) 26/29 (90%)	4e-09
Nebulin_repeat	863..891	14/29 (48%) 27/29 (93%)	2.5e-09
Nebulin_repeat	929..957	14/29 (48%) 26/29 (90%)	7.4e-07

Nebulin_repeat	965..993	6/29 (21%) 21/29 (72%)	0.0043
Nebulin_repeat	998..1026	10/29 (34%) 21/29 (72%)	0.00033
Nebulin_repeat	1033..1061	10/29 (34%) 22/29 (76%)	8.7e-05
Nebulin_repeat	1069..1097	15/29 (52%) 27/29 (93%)	1.4e-08
Nebulin_repeat	1107..1135	15/29 (52%) 26/29 (90%)	1.2e-08
Nebulin_repeat	1173..1201	11/29 (38%) 27/29 (93%)	8.9e-08
Nebulin_repeat	1209..1237	5/29 (17%) 21/29 (72%)	0.099
Nebulin_repeat	1242..1270	13/29 (45%) 25/29 (86%)	3.9e-09
Nebulin_repeat	1277..1305	12/29 (41%) 20/29 (69%)	0.00026
Nebulin_repeat	1313..1341	15/29 (52%) 27/29 (93%)	1.2e-08
Nebulin_repeat	1351..1379	14/29 (48%) 27/29 (93%)	6.7e-08
Nebulin_repeat	1417..1445	7/29 (24%) 21/29 (72%)	0.0012
Nebulin_repeat	1453..1481	6/29 (21%) 21/29 (72%)	0.13
Nebulin_repeat	1486..1514	12/29 (41%) 25/29 (86%)	2.1e-06
Nebulin_repeat	1521..1549	11/29 (38%) 23/29 (79%)	2.2e-06
Nebulin_repeat	1557..1585	16/29 (55%) 26/29 (90%)	3.1e-08
Nebulin_repeat	1595..1623	13/29 (45%) 26/29 (90%)	7.5e-08
Nebulin_repeat	1661..1689	7/29 (24%) 26/29 (90%)	2.6e-06
Nebulin_repeat	1697..1725	10/29 (34%) 24/29 (83%)	1.5e-06
Nebulin_repeat	1730..1758	10/29 (34%) 23/29 (79%)	0.00049

Nebulin_repeat	1765..1793	13/29 (45%) 26/29 (90%)	1.1e-09
Nebulin_repeat	1801..1829	12/29 (41%) 25/29 (86%)	1.3e-06
Nebulin_repeat	1839..1867	14/29 (48%) 27/29 (93%)	4.2e-09
Nebulin_repeat	1905..1933	11/29 (38%) 24/29 (83%)	1.1e-05
Nebulin_repeat	1941..1969	7/29 (24%) 19/29 (66%)	0.74
Nebulin_repeat	1974..2002	11/29 (38%) 27/29 (93%)	2.7e-08
Nebulin_repeat	2009..2037	14/29 (48%) 25/29 (86%)	3e-08
Nebulin_repeat	2045..2073	14/29 (48%) 25/29 (86%)	1e-06
Nebulin_repeat	2083..2111	13/29 (45%) 28/29 (97%)	1.8e-08
Nebulin_repeat	2149..2177	10/29 (34%) 25/29 (86%)	2.1e-05
Nebulin_repeat	2185..2213	8/29 (28%) 19/29 (66%)	0.7
Nebulin_repeat	2218..2246	11/29 (38%) 25/29 (86%)	1.4e-06
Nebulin_repeat	2253..2281	12/29 (41%) 26/29 (90%)	4.3e-10
Nebulin_repeat	2289..2317	12/29 (41%) 26/29 (90%)	5.2e-07
Nebulin_repeat	2327..2355	14/29 (48%) 28/29 (97%)	1.3e-09
Nebulin_repeat	2362..2389	12/29 (41%) 22/29 (76%)	0.49
Nebulin_repeat	2393..2421	13/29 (45%) 24/29 (83%)	7.6e-07
Nebulin_repeat	2428..2456	6/29 (21%) 21/29 (72%)	0.52
Nebulin_repeat	2461..2489	15/29 (52%) 27/29 (93%)	5.6e-10
Nebulin_repeat	2496..2524	15/29 (52%) 23/29 (79%)	4.4e-08

Nebulin_repeat	2532..2560	11/29 (38%) 25/29 (86%)	1.4e-06
Nebulin_repeat	2570..2598	13/29 (45%) 28/29 (97%)	2.5e-10
Nebulin_repeat	2636..2664	11/29 (38%) 23/29 (79%)	4e-06
Nebulin_repeat	2704..2732	11/29 (38%) 27/29 (93%)	2.2e-07
Nebulin_repeat	2739..2767	13/29 (45%) 26/29 (90%)	8.2e-09
Nebulin_repeat	2775..2803	12/29 (41%) 24/29 (83%)	5e-06
Nebulin_repeat	2813..2841	13/29 (45%) 28/29 (97%)	2.5e-10
Nebulin_repeat	2848..2875	12/29 (41%) 21/29 (72%)	0.9
Nebulin_repeat	2879..2907	10/29 (34%) 23/29 (79%)	6.2e-05
Nebulin_repeat	2914..2942	5/29 (17%) 20/29 (69%)	0.45
Nebulin_repeat	2947..2975	12/29 (41%) 27/29 (93%)	2.1e-07
Nebulin_repeat	2982..3010	13/29 (45%) 25/29 (86%)	3.9e-08
Nebulin_repeat	3018..3046	13/29 (45%) 24/29 (83%)	2.3e-06
Nebulin_repeat	3056..3084	13/29 (45%) 28/29 (97%)	2.5e-10
Nebulin_repeat	3091..3119	12/29 (41%) 21/29 (72%)	0.83
Nebulin_repeat	3122..3150	10/29 (34%) 22/29 (76%)	0.00018
Nebulin_repeat	3157..3185	8/29 (28%) 20/29 (69%)	0.077
Nebulin_repeat	3190..3218	12/29 (41%) 27/29 (93%)	2.1e-07
Nebulin_repeat	3225..3253	12/29 (41%) 26/29 (90%)	6.5e-08
Nebulin_repeat	3261..3289	14/29 (48%) 26/29 (90%)	6.7e-08

Nebulin_repeat	3299..3327	14/29 (48%) 25/29 (86%)	2.5e-06
Nebulin_repeat	3365..3393	14/29 (48%) 25/29 (86%)	1.8e-08
Nebulin_repeat	3400..3428	11/29 (38%) 21/29 (72%)	5.9e-05
Nebulin_repeat	3433..3461	13/29 (45%) 27/29 (93%)	1.1e-08
Nebulin_repeat	3468..3496	15/29 (52%) 26/29 (90%)	5.5e-09
Nebulin_repeat	3504..3532	13/29 (45%) 26/29 (90%)	1.8e-07
Nebulin_repeat	3542..3570	12/29 (41%) 28/29 (97%)	3.9e-09
Nebulin_repeat	3608..3636	12/29 (41%) 22/29 (76%)	0.00011
Nebulin_repeat	3643..3671	7/29 (24%) 20/29 (69%)	0.07
Nebulin_repeat	3676..3704	15/29 (52%) 25/29 (86%)	1e-08
Nebulin_repeat	3711..3739	11/29 (38%) 27/29 (93%)	1.5e-08
Nebulin_repeat	3747..3775	15/29 (52%) 26/29 (90%)	1.6e-07
Nebulin_repeat	3785..3813	11/29 (38%) 25/29 (86%)	2.7e-07
Nebulin_repeat	3851..3879	14/29 (48%) 24/29 (83%)	4.8e-07
Nebulin_repeat	3919..3947	16/29 (55%) 25/29 (86%)	3.6e-09
Nebulin_repeat	3954..3982	13/29 (45%) 26/29 (90%)	5.3e-09
Nebulin_repeat	3989..4017	14/29 (48%) 27/29 (93%)	1.1e-08
Nebulin_repeat	4027..4055	10/29 (34%) 23/29 (79%)	0.00011
Nebulin_repeat	4093..4121	13/29 (45%) 24/29 (83%)	6.8e-07
Nebulin_repeat	4128..4156	8/29 (28%) 19/29 (66%)	0.15

Nebulin_repeat	4161..4189	9/29 (31%) 24/29 (83%)	5.2e-07
Nebulin_repeat	4195..4223	9/29 (31%) 25/29 (86%)	6.1e-06
Nebulin_repeat	4231..4259	11/29 (38%) 24/29 (83%)	2.6e-06
Nebulin_repeat	4269..4297	11/29 (38%) 27/29 (93%)	5.6e-08
Nebulin_repeat	4335..4363	11/29 (38%) 24/29 (83%)	5.9e-06
Nebulin_repeat	4370..4398	8/29 (28%) 23/29 (79%)	0.0069
Nebulin_repeat	4405..4433	20/29 (69%) 25/29 (86%)	1.2e-10
Nebulin_repeat	4440..4468	12/29 (41%) 23/29 (79%)	8.3e-06
Nebulin_repeat	4476..4504	14/29 (48%) 21/29 (72%)	0.00042
Nebulin_repeat	4549..4577	8/29 (28%) 21/29 (72%)	0.14
Nebulin_repeat	4580..4608	8/29 (28%) 24/29 (83%)	2e-05
Nebulin_repeat	4615..4643	11/29 (38%) 26/29 (90%)	1.8e-06
Nebulin_repeat	4650..4678	13/29 (45%) 23/29 (79%)	2.1e-06
Nebulin_repeat	4685..4713	7/29 (24%) 23/29 (79%)	0.00067
Nebulin_repeat	4721..4749	12/29 (41%) 23/29 (79%)	2.8e-05
Nebulin_repeat	4759..4787	10/29 (34%) 20/29 (69%)	0.015
Nebulin_repeat	4794..4822	8/29 (28%) 21/29 (72%)	0.14
Nebulin_repeat	4825..4853	9/29 (31%) 20/29 (69%)	0.037
Nebulin_repeat	4860..4888	13/29 (45%) 23/29 (79%)	2.3e-06
Nebulin_repeat	4895..4923	12/29 (41%) 21/29 (72%)	0.00025

Nebulin_repeat	4966..4994	12/29 (41%) 20/29 (69%)	0.0011
Nebulin_repeat	5002..5030	7/29 (24%) 20/29 (69%)	0.37
Nebulin_repeat	5037..5065	12/29 (41%) 22/29 (76%)	3.3e-05
Nebulin_repeat	5072..5100	11/29 (38%) 22/29 (76%)	0.0029
Nebulin_repeat	5107..5135	8/29 (28%) 22/29 (76%)	0.024
Nebulin_repeat	5142..5170	8/29 (28%) 23/29 (79%)	0.00063
Nebulin_repeat	5177..5205	13/29 (45%) 22/29 (76%)	0.2
Nebulin_repeat	5212..5240	11/29 (38%) 21/29 (72%)	0.028
Nebulin_repeat	5247..5275	13/29 (45%) 23/29 (79%)	3.1e-08
Nebulin_repeat	5282..5310	11/29 (38%) 24/29 (83%)	0.21
Nebulin_repeat	5317..5345	13/29 (45%) 23/29 (79%)	3e-06
Nebulin_repeat	5352..5380	15/29 (52%) 24/29 (83%)	2.1e-07
Nebulin_repeat	5387..5415	10/29 (34%) 22/29 (76%)	2.1e-05
Nebulin_repeat	5422..5450	12/29 (41%) 24/29 (83%)	7.9e-06
Nebulin_repeat	5457..5485	13/29 (45%) 24/29 (83%)	1.9e-07
Nebulin_repeat	5493..5521	12/29 (41%) 24/29 (83%)	2.1e-06
Nebulin_repeat	5563..5591	13/29 (45%) 21/29 (72%)	0.00026
Nebulin_repeat	5598..5626	14/29 (48%) 22/29 (76%)	0.0018
Nebulin_repeat	5633..5661	13/29 (45%) 23/29 (79%)	1.3e-05
Nebulin_repeat	5670..5698	12/29 (41%) 25/29 (86%)	4.7e-06

Nebulin_repeat	5707..5735	11/29 (38%) 21/29 (72%)	0.0097
Nebulin_repeat	5741..5769	11/29 (38%) 22/29 (76%)	0.00015
Nebulin_repeat	5776..5804	17/29 (59%) 25/29 (86%)	2.2e-08
Nebulin_repeat	5848..5876	9/29 (31%) 24/29 (83%)	0.00049
Nebulin_repeat	5883..5911	16/29 (55%) 25/29 (86%)	4e-07
Nebulin_repeat	5918..5946	10/29 (34%) 27/29 (93%)	1.1e-06
Nebulin_repeat	5955..5983	8/29 (28%) 25/29 (86%)	1.1e-05
Nebulin_repeat	5992..6020	10/29 (34%) 26/29 (90%)	1.5e-07
Nebulin_repeat	6030..6058	10/29 (34%) 26/29 (90%)	1e-07
Nebulin_repeat	6065..6093	15/29 (52%) 26/29 (90%)	1.3e-08
Nebulin_repeat	6100..6128	12/29 (41%) 26/29 (90%)	5e-06
Nebulin_repeat	6135..6163	14/29 (48%) 22/29 (76%)	5.7e-06
Nebulin_repeat	6166..6194	10/29 (34%) 20/29 (69%)	0.066
Nebulin_repeat	6197..6225	10/29 (34%) 20/29 (69%)	0.048
Nebulin_repeat	6228..6256	12/29 (41%) 22/29 (76%)	4e-05
Nebulin_repeat	6259..6287	11/29 (38%) 22/29 (76%)	1.1e-05
Nebulin_repeat	6290..6318	11/29 (38%) 22/29 (76%)	1.1e-05
Nebulin_repeat	6321..6349	11/29 (38%) 24/29 (83%)	5.5e-07
Nebulin_repeat	6352..6380	11/29 (38%) 22/29 (76%)	9.6e-06
Nebulin_repeat	6383..6411	11/29 (38%) 22/29 (76%)	6.7e-05

Nebulin_repeat	6414..6442	10/29 (34%) 21/29 (72%)	2e-05
Nebulin_repeat	6450..6478	14/29 (48%) 25/29 (86%)	3e-09
SH3	6644..6700	24/58 (41%) 47/58 (81%)	2.6e-17

Example 36.

The NOV36 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 36A.

5

Table 36A. NOV36 Sequence Analysis			
	SEQ ID NO: 145	2973 bp	
NOV36a, CG120166-01 DNA Sequence	ATGTCGGAAGAAACCCGACAGAGCAAATTGGCCGAGCGAAGAAAAAGTTGAGAGAATATCAG CAGAGGAATAGCCCTGGTGTTCCTACAGGAGCGAAAAAGAAGAAAAATAAAAAATGGCAGT AACCCCTGAGACAACCACTTCTGGTGGTTGCCACTCACCTGAGGATACACCAAGGACAATGCT GCTACTCTACAACCATCTGATGACACCGTGTACCTGGCGGTGCCCTTCCCTGGTGCCAGT CTCACTAGCATGGCGGCATCTCAGAATCATGATGCTGACAATGTCCCTAATCTCATGGATGAA ACCAAGACTTTCTCATCAACCGAGAGCCTGCGACAACCTCCCAACAGCTCAATGGTCTGTGTT GTGAGTCTGCGACATGTGTCAATGGGGAGGGCCCTGCATCGTCTGCTAACCTGAAGGATCTG GAGAGCCGGTACCAACAGCTAGCCGTAGCCCTGGACTCCAGCTATGTAACAAACAAACAACCTC AATATCACGATAGAGAAATTGAAACAACAGAACCAAGAAATTACGGATCAGTTGGAAGAAGAA AAGAAAGAATGCCACCAAAAGCAGGGAGCCCTAAGGGAGCAGTTACAGGTTACATTTCAGACC ATAGGGATCCTCGTATCAGAGAAAGCTGAGTTACAGACAGCCCTGGCTCACACTCAGCATGCT GCCAGGCAGAAAGAAGGAGAGTCTGAAGATCTGGCCAGCCGCTGCAGTATTCCTGGCGGCGT GTGGGAGAGTTGGAGCGGCTCTCTGTGTCTCCACGCAGCAGAAGAAGCAGACAGGTAC AACAAGGAGTTAACCAGAGAGAGACGCCCTCAGGCTGGAGTTATACAAGAACACCCAAAGC AATGAGGACCTGAAGCAAGAGAATCAGAATTGGAAGAGAAGCTTCGGGTCCTAGTGACTGAG AAGGCTGGCATGCAGCTTAACCTGGAAGAATTGCAAAAGAAGTTAGAGATGACGGAACCTCTG CTTCAACAGTTTTCAAGCCGGTGTGAAGCCCTGATGCTAACCAGCAGTTACAGCAGGCCATG GAGGAGCGGCGACAGCTGGAAGCACACCTGGGGCAGGTAATGGAGTCGGTTAGACAACTACAA ATGGAGAGAGATAAATATGCGGAGAATCTCAAAGGAGAGAGCGCCATGTGGCGGCAGAGGATG CAGCAGATGTGAGAGCAGGTGCACACATTGAGAGAGGAGAAGGAATGTAGCATGAGTCGGGTA CAGGAGCTGGAGACGAGCTTGGCTGAACCTGAGGAACAGATGGCTGAACCCCGCCCCAGAG CCCCCAGCAGGGCCCTCGAGGTGGAGCAGCAGCTACAAGCGGAGGCTGAGCACCTGCGGAAG GAGCTGGAGGCTCTGGCAGGACAGCTTCAAGCCAGGTGCAAGACAATGAGGGCTTGAGTCGC CTGAACCGGGAGCAGGAGGAGAGGCTGCTGGAGCTGGAGCGGGCGGCCGAGCTCTGGGGGAG CAGGCGGAGGCGCGCAGGCAATCTGGAGACCATGCAGAACGACCGCACTACCATCAGCCGC GCACTCTCCAGAACCGGAGCTCAAGGAGCAGCTGGCTGAGCTGCAGAGCGGATTTGTAAG CTGACTAATGAGAATGAGATCACCAGCGCACTGCAGTCGGAGCAGCAGCTCAAGAGGGAG CTGGGAAAGAAGCTGGGCGAGCTGCAGGAGAAGCTGAGCGAGCTGAAGGAAACGGTGGAGCTG AAGAGCCAAGAGGCTCAAAGTCTGCAGCAGCAGCGAGACCACTGGGACACCTGCAGCAG TATGTGGCCGCTATCAGCAGCTGACCTCTGAGAAGGAGGTGCTGCATAATCAGTCTACTGCTG CAGACCCAGCTCGTGGACAGCTGCAGCAGCAGGAAGCTCAGGGCAAGCGGTGGCCGAGATG GCCCGCAAGAGTTGCAGGAAACCCAGGAGCGCCTGGAAGCTGCCACCCAGCAGAATCAGCAG CTACGGGCCAGTTGAGCCTCATGGCTCACCTGGGGAAGGAGATGGACTGGACCGGGAGGAG GAGGAGGATGAGGAGGAGGAGGAGGAGGAGGCGGTGGCAGTACCTCAGCCCATGCCAAGCATC CCGGAGGACCTGGAGAGCCGGAAGCCATGGTGGCATTTCCTCACTCAGCTGTAGCCAGTGCC GAGGAGGAGCAGGCAAGGCTACGTGGGAGCTGAAGGAGCAAGGGTGGCTGCGCTGCCGCGCTG GCTCACCTGCTGGCCTCGGCCAGAAAGGAGCCTGAGGCAGCAGCCCCAGCCCCAGGGACCGGG GGTGATTCTGTGTGTGGGAGACCCACCGGGCCCTGCAGGGGGCATGGAGAAGCTGCAGAGC CGCTTTATGAGCTCATGTCAGGAGAAGGCAGACCTGAAGGAGAGGGTAGAGGAAGTGAACAT CGTGCATCCAGCTTTCTGGAGAGACAGACCATTTGGAGAGTACATTGCACTGTACCAAGCAGC CAGAGGGCAGTGCTGAAGGAGCGGCACCGGAGAAGGAGGAGTACATCAGCAGGCTGGCCCCA GACAAGGAGGAGATGAAGGTGAAGCTGTGGAGCTGCAGGAGCTGGTCTTACGGCTGTGGGC GACCGCAACGAGTGGCATGGCAGATTCTGGCAGCTGCCAGAACCTGTGATGAGCCACT		

	TCAGGGGCCCCAGCCCCCAGGAACCTGGGGCTGCCAACAGCAGGGTGATCTTTGCGAGGTG AGCCTCGCCGCGCAGTGTGGAGCCTGCCCAAGGAGAGGCCAGGGAGGGTTCTCCCGTGACAAC CCCCTGCACAGCAGATCATGCAGCTGCTTCGTGAGATGCAGAACCCCGGGAGCGCCAGGC TTGGGCAGCAACCCCTGCATTCTTTTTTTTACCGGGCTGACGAGAATGATGAGGTGAAGATC ACTGTCATCTAA		
	ORF Start: ATG at 1	ORF Stop: TAA at 2971	
	SEQ ID NO: 146	990 aa	MW at 111657.5kD
NOV36a, CG120166-01 Protein Sequence	MSEETRQSKLAAAKKLREYQQRNSPGVPTGAKKKKKIKNGSNPETTTSGGCHSPEDTPKDNA ATLQPSDDTVLPGGVPSPGASLTSMASQNHADNPNLMDETKFSSTESLRQLSQQLNGLV CESATCVNGEGPASSANLKDLESRYQLAVALDSSVYTNKQLNITIEKLKQONQEI TDQLEEE KKECHQKQKALREQLQVHIQTIGILVSEKAEQLTALAHTQHAARQKEGESEDLASRLQYSRRR VGELERALS AVSTQKKADRYNKELTKERDALRLLEYKNTQSNEDLKQEKSELEEKLRVLVTE KAGMQLNLEELQKKLEMTLELLQFSSRCEAPDANQQLQOAMEERAQLEAHLGQVMESVRQLQ MERDKYAENLKGESAMWRQRMQMSQVHTLREEKECSMSRVQLETLAELRNQMAEPPPPPE PPAGPSEVEQQLQAEAEHLRKELEGLAGQLQAVQDNEGLSRLNREQEERLLELERAELWGE QAEARRQILETMQNDRTTISRALSQNRLEKEQLAELQSGFVKLTNENMEITSALQSEQHVKRE LGKKLGELQEKLESELKETVELKSQEAQSLQQQORDQYLGHLLQYVAAAYQQLTSEKEVLHNQLLL QTQLVDQLQQEAQKGAVAMARQELQETQERLEAATQONQLRAQLSLMAHPGEGDGLDREE EEDDEEEEEEA VAVPQPMPSIPEDLESREAMVAFFNSAVASAEQARLRGQLKEQVRRCRRRL AHLASAQKEPEAAAPAGPTGGDSVCGETHRALQGAMEKLSRFMELMQEKADLKERVELEH RCIQLSGETDTIGEYIALYQSQR AVLKERHREKEEYISRLAQDK EEMKVKLLELQELVLRVLV DRNEWHGRFLAAAQNPADPTSGAPAPQELGAANQQGDLCEVSLAGSVEPAQGEAREGSPRDN PTAQQIMQLLREMQRPRERPLGSGNPCIPFFYRADENDEVKITVI		
	SEQ ID NO: 147	2886 bp	
NOV36b, CG120166-02 DNA Sequence	CCTCCCCCGCCCGCGATGTCGGAAGAAACCCGACAGAGCAAATTGGCCGCGAGCGAAGAAAAAG TTGAGAGAATATCAGCAGAGGAATAGCCCTGGTGTCTTCTACAGGAGCGAAAAAGAGAAGAAA ATAAAAAATGGCAGTAACCTGAGACAACCACTTCTGGTGGTTGCCACTCACCTGAGGATACA CCCAAGGACAATGCTGCTACTCTACAACCATCTGATGACACCGTGTTACCTGGCGGTGTCCCT TCCCTGGTGCCAGTCTCACTAGCATGGCGGCATCTCAGAATCATGATGCTGACAATGTCCCT AATCTCATGGATGAAACCAAGACTTTCTCATCAACCGAGAGCCTGCGACAACCTCTCCCAACAG CTCAATGGTCTTGTGTGAGTCTGCGACATGTGTCAATGGGGAGGGCCCTGCATCGTCTGCT AACCTGAAGGATCTGGAGAGCCGTACCAACAGCTAGCGGTAGCCCTGGACTCCAGCTCAGTATGTA ACAAACAACAACCTCAATATCAGATAGAGAAATTGAAACAACAGAACCAAGAAATTACGGAT CAGTTGGAAGAAGAAAAGAAAGAAATGCCACCAAAAGCAGGGAGCCCTAAGGGAGCAGTTACAG GTTACATTCAGACCATAGGGATCCTCGTATCAGAGAAAGCTGAGTTACAGACAGCCCTGGCT CACACTCAGCATGCTGCCAGGCAGAAAGAAGGAGAGTCTGAAGATCTGGCCAGCCGCTGCAG TATTCCCGCGCGCGTGTGGGAGAGTTGGAGCGGGCTCTCTCTGCTGTCTCCACGCAGCAGAAG AAGGCAGACAGGTACAACAAGGAGTTAACCAGAGAGAGAGACGCCCTCAGGCTGGAGTTATAC AAGAACAACCAAGCAATGAGGACCTGAAGCAAGAGAAATCAGAATTGGAAGAGAAGCTTCGG GTCTTAGTGACTGAGAAGGCTGGCATGCAGCTTAACCTGGAAGAATTGCAAAAGAAGTTAGAG ATGACGGAACCTCTCGTTCACAGTTTTCAAGCCGGTGTGAAGCCCTGATGCTAACACAGCAG TTACAGCAGGCCATGGAGGAGCGGGCAGAGCTGGAAGCACACCTGGGGCAGGTAATGGAGTCG GTTAGACAACCTACAAATGGAGAGAGATAAATATGCGGAGAATCTCAAAGGAGAGAGCGCCATG TGGCGGCAGAGGATGCAGCAGATGTGAGAGCAGGTGCACACATTGAGAGAGGAGAGAAGGAATGT AGCATGAGTCGGGTACAGGAGCTGGAGACGAGCTTGGCTGAACCTGAGGAACCATGGCTGAA CCCCCGCCCCAGAGCCCCAGCAGGGCCCTCCGAGGTGGAGCAGCAGCTACAAGCGGAGGCT GAGCACCTGCGGAAGGAGCTGGAGGGTCTGGCAGGACAGCTTCAAGCCCAGGTGCAAGACAAT GAGGGCTTGAGTCGCCTGAACCGGGAGCAGGAGGAGAGGCTGCTGGAGCTGGAGCGGGCGGCC GAGCTCTGGGGGAGCAGCGGAGCGGAGCGCAGGCAAACTCTGGAGACCATGCAGAACGACCGC ACTACCATCAGCCGCGCACTCTCCAGAACCGGGAGCTCAAGGAGCAGCTGGCTGAGCTGCAG AGCGGATTGTAAAGCTGACTAATGAGAACATGGAGATCACCAGCGCACTGCAGTCGGAGCAG CAGCTCAAGAGGGAGCTGGGAAAGAAGCTGGGCGAGCTGCAGGAGAAGCTGAGCGAGCTGAAG GAAACGGTGGAGCTGAAGAGCCAAGAGGCTCAAAGTCTGCAGACCCAGCTCGTGAGCCAGCTG CAGCAGCAGGAAGCTCAGGGCAAAGCGGTGGCCGAGATGGCCCGCAAGAGTTGCAAGAAACC CAGGAGCGCTGGAAGCTGCCACCCAGCAGAATCAGCAGCTACGGGCCCAGTTGAGCCTCATG GCTCACCTTGGGGAAGGAGATGGACTGGACCGGAGGAGGAGGAGGATGAGGAGGAGGAGGAG GAGGAGGCGGTGGCAGTACCTCAGCCCATGCCAAGCATCCGGAGAGCATGGAGAGCCGGGAA GCCATGGTGGCATTTTCAACTCAGCTGTAGCCAGTGCCGAGGAGGAGCAGGCAAGCTACGT GGGCAGCTGAAGGAGCAAAGGGTGCCTGCCGGCGCTGGCTCACCTGCTGGCCTCGGCCAG AAGGAGCCTGAGGCAGCAGCCCCAGCCCCAGGGACCGGGGTGATTCTGTGTGTGGGGAGACC CACCGGGCCCTGCAGGGGGCCATGGAGAAGCTGCAGAGCCGCTTTATGGAGCTCATGCAGGAG AAGGCAGACCTGAAGGAGAGGGTAGAGGAACCTGCAACATCGCTGCATCCAGCTTTCTGGAGAG ACAGACACCATTTGAGAGTACATTGCACTGTACCAGAGCCAGAGGCGAGTCTGAAGGAGCGG CACCGGGAGAAGGAGGAGTACATCAGCAGGCTGGCCCAAGACAAGGAGGAGATGAAGGTGAAG		

	CTGCTGGAGCTGCAGGAGCTGGTCTTACGGCTTGTGGGCGACCGCAACGAGTGGCATGGCAGA TTCTGGCAGCTGCCAGAACCTGCTGATGAGCCACTTCAGGGGCCCCAGCCCCCAGGAA CTTGGGGCTGCCAACAGCAGGGTGATCTTTGCGAGGTGAGCCTCGCCGGCAGTGTGGAGCCT GCCCCAAGGAGAGGCCAGGGAGGGTTCTCCCGTGACAACCCCACTGCACAGCAGATCATGCAG CTGCTTCGTGAGATGCAGAACCCCGGGAGCGCCAGGCTTGGGCAGCAACCCCTGCATTCTT TTTTTTACCGGGCTGACGAGAATGATGAGGTGAAGATCACTGTCATCTAA		
	ORF Start: ATG at 16		ORF Stop: TAA at 2884
	SEQ ID NO: 148	956 aa	MW at 107591.0kD
NOV36b, CG120166-02 Protein Sequence	MSEETRQSKLAAAKKKLREYQQRNSPGVPTGAKKKKKIKNGSNPETTTSGGCHSPEDTPKDNA ATLQPSDDTVLPGGVPSPGASLTSMASQNHADNPNLMDETKTFSSTESLRQLSQQLNGLV CESATCVNGEGPASSANLKDLESRYQLAVALDSSVYTNKQLNITIEKLQQNQEIITDQLEEE KKECHQKQGALREQLVHIQTIGILVSEKAELQTALAHTQHAAHQKEGESEDLASRLQYSRRR VGELERALS AVSTQKKADRYNKELTKERDALRLELYKNTQSNEDLKQEKSELEEKLRVLVTE KAGMQLNLEELQKKLEMTLELLQQFSSRCEAPDANQQLQQAEMEERAQLEAHLGQVMESVRQLQ MERDKYAENLKGESAMWRQRMQMSQVHTLREEKECSMSRVQLETS LAELRNQMAEPPPE PPAGPSEVEQQLQAEAEHLRKELEGLAGQLQAVQDNEGLSRLNREQEERLLELERAELWGE QAEARRQILETMQNDRTTISRALSQNRELKEQLAELQSGFVKLTNENMEITSALQSEQHVKRE LGKKLGELQEKLSSELKETVELKSQEAQSLQTLVDQLQQEAGQKAVAEMARQELQETQERLE AATQQNQQLRAQLSLMAHPGEGDGLDREEEDEEEEEEA VAVPQPMPSIPEDLESREAMVAF FNSAVASAEQEARLQRLKEQVRRCRRLLASQAQKEPEAAAPAGTGGDSVCGETHRALQ GAMEKLQSRFMELMQEKADLKERVEELEHRCIQLSGETDTTIGEYIALYQSQRVFLKERHREKE EYISRLAQDKEEMKVKLLELQELVLRVLDNRNEWHGRFLAAQNPADPTSGAPAPQELGAAN QQGDLCEVSLAGSVEPAQGEAREGSPRDNPTAQQIMQLLREMQRPRRPGLSGNPCIPFFYRA DENDEVKITVI		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 36B.

Table 36B. Comparison of NOV36a against NOV36b.		
Protein Sequence	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV36b	1..990	790/990 (79%)
	1..956	790/990 (79%)

5

Further analysis of the NOV36a protein yielded the following properties shown in Table 36C.

Table 36C. Protein Sequence Properties NOV36a	
PSort analysis:	0.6000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV36a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 36D.

Table 36D. Geneseq Results for NOV36a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM93868	Human polypeptide, SEQ ID NO: 3973 - Homo sapiens, 360 aa. [EP1130094-A2, 05-SEP-2001]	88..438 1..351	348/351 (99%) 348/351 (99%)	0.0
ABG20674	Novel human diagnostic protein #20665 - Homo sapiens, 1262 aa. [WO200175067-A2, 11-OCT-2001]	148..960 578..1261	360/842 (42%) 466/842 (54%)	e-146
ABG20672	Novel human diagnostic protein #20663 - Homo sapiens, 1717 aa. [WO200175067-A2, 11-OCT-2001]	128..840 1038..1632	269/738 (36%) 364/738 (48%)	6e-89
ABG03395	Novel human diagnostic protein #3386 - Homo sapiens, 633 aa. [WO200175067-A2, 11-OCT-2001]	294..633 55..403	181/367 (49%) 228/367 (61%)	2e-78
ABG20896	Novel human diagnostic protein #20887 - Homo sapiens, 1213 aa. [WO200175067-A2, 11-OCT-2001]	503..831 332..635	155/332 (46%) 199/332 (59%)	6e-62

In a BLAST search of public sequence databases, the NOV36a protein was found to have homology to the proteins shown in the BLASTP data in Table 36E.

5

Table 36E. Public BLASTP Results for NOV36a				
Protein Accession Number	Protein/Organism/Length	NOV36a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NYF9	Golgi matrix protein GM130 - Homo sapiens (Human), 990 aa.	1..990 1..990	990/990 (100%) 990/990 (100%)	0.0
Q62839	Cis-golgi matrix protein GM130 - Rattus norvegicus (Rat), 986 aa.	1..990 1..986	748/1003 (74%) 841/1003 (83%)	0.0
Q08379	Golgin-95 - Homo sapiens (Human), 620 aa.	371..990 1..620	620/620 (100%) 620/620 (100%)	0.0
Q921M4	Unknown (protein for MGC:11816) - Mus musculus (Mouse), 617 aa.	371..990 1..617	466/627 (74%) 522/627 (82%)	0.0
Q9BRB0	Hypothetical 38.8 kDa protein - Homo sapiens (Human), 345 aa.	516..813 1..297	296/298 (99%) 296/298 (99%)	e-160

PFam analysis predicts that the NOV36a protein contains the domains shown in the Table 36F.

Table 36F. Domain Analysis of NOV36a

Pfam Domain	NOV36a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5 **Example 37.**

The NOV37 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 37A.

Table 37A. NOV37 Sequence Analysis

	SEQ ID NO: 149	2490 bp	
NOV37a, CGI20401-01 DNA Sequence	AAAAGGGCATTGAGACCNCTGGGAGGTGACTTCATCAATTGCTCNTACCCCGGAGTTCTCTAGAC AGACTTGGACNATCCCTACCTTGGGTGTGCATCTCTTGGGAAACGGAGGGGCCNAGAGGTT CGAATCTCNTCTTCAGCCTTTTAAAGCTTCACTTGGTCAGAATCCTTGGATGAGCNCTGTGGGAC CGTTCTCTCATAGCCCGGTGGTTTGANCCAGTGGCTTTGGGACTGTAAGAGGATGGACAAAGGT AGTATGAGGCCGGGACAGCTCCCCAAATGCTCGGGCGCCAAGCATCTGCTGATTCCACTCTGT GCCTGGCAGCTGTGTCTTTCCGTCTCTGGCTAGGCCTTGCCACTCCCCCTGGGCGCAGCACCA CCATGTACCTGTCTGGCAGAGCCACAGTTGACTGACTCTGTTGACTGCTGCTCTTTAGACAT GGCTGCATTTTGATGGCAGAGGGAGCAGTTATGCCCAGAGCCCAGCGTGTCTGGCATCACCAGT CTGACCTGAGCACCATTGTGCTGCACGACATGACACGTGGCACCGGGGGAAGTCCGACGCTGG AAGGTCTGGGCCAGGTCTGAGCCAGATGGGATCTGGATGGCCAAGGAACCTTACCTTAAAC CTCAAGTGTAAGGAGCAGGGGAGGGGCCAAGAGGGCTGGCAGGGGAAGGGGGCTGGGGAGG GGTCCCATTTGTCAGAGGCTCTTAGGATCTTAGGAGGCCCGAATCCCACCATTTCTCTACTTGG TAGATCTCAGGGGCTGCTAGATTATCCCTGATGGCATCAGGCACTGCCAGCCGCTCAGAGGA TGAGGAGTCACTGGCAGGGCAGAAGCGAGCCCTCTCCAGGCCTTGGGCACCATTCCCTAAACG GAGAAGCTCTCTCAGGTTTCATCAGAGGAAGAAGTTCGATGATGAGCTGGTGGAGAGAGCCT GGCAAAATCTTCTACCCGGGCAAAGGGGGCCAGTGGGTGGAACAGGCGCTGTTCGGGGAG TGAACCCCTCTCCAGTGAGAAGAAGAAGGTATCCAAGCCCCCAGCACTCTGTGCCACCCAG CCCAGCCCCAGCCCCCTGGACTCACCAGCGTGTGAAGAAGAGTTAAACAGCCACTTCAGGTGAC CAAGGATCTGGGCCGCTGGAAGCCTGCAGATGACCTCTGCTCATAAATGCTGTGTTGTCAGAC CAACGACCTGACCTCCGTCCACTGGCGTGAATTCAGTGC CGCTTCAACCTTCCGGAGGT CCAGGAGCGTTGGTACGCCCTGCTCTACGATCCTGTCTATCTCCAAGTTGGCCTGTCAGGCCAT GAGGCAGCTGCACCCAGAGGCTATTGACGCCATCCAGAGCAAGGCCCTGTTTAGCAAGGCTGA GGAGCAGCTGCTGAGCAAAGTGGGATCGACCAAGCCAGCCACCTTGGAGACCTTCCAGGACCT GCTGCACAGACACCTGATGCCTTCTACCTGGCCCGTACCGCGAAGGCCCTGCAGGCCCTGCT GCAGCTCATGAAGCAGATTATTAACCTGCTGGAGGACAGACAGTGCAGCCGCTGCCCAAAGGGA CCAAGTGCTGAATTTCTCTGATGCAGAGGACCTGATTGATGACAGTAAGCTCAAGGACATGCG AGATGAGGTCTGGAACATGAGCTGATGGTGGCTGACCGGCCGAGAACGAGAGATTCCGCA GCTGGAACAGGAACCTGCATAAGTGGCAGGTGCTAGTGGACAGCATCACAGGCATGAGCTCTCC GGACTTCGACAACAGACACTGGCAGTGTGCGGGGCCGATGGTGCAGTACCTGATCGGCTCT GCGTGAGATCACCTGGGCAGAGCAACCAAGGATAAACCAGATTGATGTGGACCTGTCTCTGGA GGGTCCGGCTTGAAGATATCCCGAAACAAGGTGTCATCAAGCTGAAGAACAACGGTGATTT CTTCAATTGCCAATGAGGGTCGACGGCCCATCTACATCGATGGACGGCCGGTGTCTGTGGCTC CAAATGGCCCTCAGCAACAACCTCTGTGGTGGAGATCGCCAGCCTGCGATTCTCTCTTAT CAACGAGGACCTCATCTGAGGCTGAGGCTGCCAAGATCACACCAAGTGAAGGAGT GGTGGCAGGACTCGTGGGCCCTCTCCGGCCTGTTTCCCTGCCACTCCAGCCCCCTTGAGCTG GGAACCTCAGGCTCTTGGAAAACCTGGGCAGTGGGAGGCTCAGCTGCGGGCCATTGATTGAG CCTTGTAGGGAGGATAGGGCTGGCCTTTGTGAAGCCAGCAGAGGCTGAGAACCTCAGGCTTCC CTATGACAGAGCCCTCCCATCTCTCTCTTAAAAAACAACCTACCCCCATTGCCACC TTCACTCTGTGCTCTCCAGCTGATTAGCCTCAGACTCTCTTTTATTGTTTTCTTTTGTAA		

	TAAAAAGCACCAGGTTCAAAAAA		
	ORF Start: ATG at 533		ORF Stop: TGA at 2135
	SEQ ID NO: 150	534 aa	MW at 59016.8kD
NOV37a, CG120401-01 Protein Sequence	MTRGTGGTAQGRSGPGLSPDGIWMAKELYLKTSSVKEAGEGPRGLAGEGGWGGVPFAEALRI LGGPNPTISLLARSQGLDSSLMASGTASRSEDEESLAGQKRASSQALGTIPKRSSSRFIKR KKFDELVESSLAKSSTRAGASGVEPGRCSGSEPSSEKKKVSAPSTPVPPSPAPAPGLTK RVKKSQKPLQVTKDLGRWKPADDLLLINAVLQTNDLTSVHLGVKFSCRFTLREVQERWYALLY DPVISKALACQAMRQLHPEAIAAIQSKALFSKAEQLLSKVGSTSQPTLETQDLLHRHPDAFY LARTAKALQAHWQLMKQYLLLEDQTVQPLPKGDQVLNFSDAEDLIDDSKLDKMRDEVLEHELM VADRRQKREIRQLEQLHKWQVLVDSITGMSPPDFDNQTLAVLRGRMVRYLMRSREITLGRAT KDNQIDVDLSLEGPAPWKISRQGVIKLKNNGDFFIANEGRRPIYIDGRPVLCGSKWRLSNNSV VEIASLRFVFLINQDLIALIRAEAAKITPQ		
	SEQ ID NO: 151	1764 bp	
NOV37b, CG120401-02 DNA Sequence	TACCCCGAGAGTTCTTAGACAGACTTGGACCGATCCCCTACCTTGGGTGTGCACTCTTGGGAG AACGAGGGGCGGAGGGTCCGAATCTCCTCTCAGCCTTTAAGCTCACCTGGTCAGAATCCTTGG ATGAGCCTGTGGGACCGTTCTCTAGCCCGGTGGTTTGGAAACAGTGGCTTTGGGACTGTAA GAGGATGGACAAAGATTCTCAGGGGTGCTAGATTATCCCTGATGGCATCAGGCATCGCCAG CCGCTCAGAGGATGAGGAGTCACTGGCAGGGCAGAAGCGAGCCTCTCCAGGCCTTGGGCAC CATCCCTAAACGGAGAAGTCTCTCAGGTTTCAAGAGGAAGAAGTTCGATGATGAGCTGGT GGAGAGCAGCCTGGCAAAATCTTCTACCCGGGCAAAGGGGGCCAGTGGGGTGGAAACAGGGCG CTGTTCCGGGAGTGAACCTCTCCAGTGAGAAGAAGAAGGTATCCAAAGCCCCAGCACTCC TGTGCCACCCAGCCCGCCCGCTGGACTCACCAGCGTGTGAAGAAGTAAACAGCC ACTTCAGGTGACCAAGGATCTGGGCCGCTGGAAGCCTGCAGATGACCTCTGCTCATAAATGC TGTGTTGCAGACCAACGACCTGACCTCCGTCCACCTGGGCGTGAAACTCAGCTGCCGCTTAC CCTTCGAGAGGTCCAGGAGCGTTGGTACGCCCTGCTCTACGATCTGTCTCTCAAGTTGGC CTGTGAGCCATGAGGCAGCTGCACCCAGAGGCTATTGCAAGCCATCCAGAGCAAGGCCCTGCA GGCCCACTGGCAGCTCATGAAGCAGTATTACCTGCTGGAGGACCAGACAGTGCAGCCGCTGCC CAAAGGGGACCAAGTGTGAACCTTCTCTGATGCAGAGGACCTGATTGATGACAGTAAGCTCAA GGACATGCGAGATGAGGTCTTGAACATGAGCTGATGGTGGCTGACCGGCGCCAGAAGCGAGA GATTGCGGACGTGGAACAGGAAGTGCATAAGTGGCAGGTGCTAGTGGACAGCATCACAGGCAT GAGCTCTCCGACTTCGACAACAGACACTGGCAGTACTGCGGGGCGCATGGTGGCGGTACCT GATGCGCTCGCGTGAGATCACCCCTGGGCAGAGCAACCAAGGATAACCAGATTGATGTGGACCT GTCTCTGGAGGGTCCGGCCTGGAAGATATCCCGAAACAAGGTGTCTACAGCTGAAGAACAA CGGTGATTTCTTCATTGCCAATGAGGGTGCAGCGCCCATCTACATCGATGGACGGCCGGTGCT CTGTGGCTCCAAATGGCGCCTCAGCAACAACCTCTGTGGTGGAGATCGCCAGCCCTGCGATTCT TCTCCTTATCAACCAGGACCTCATTGCCCTCATCAGGCTGAGGCTGCCAAGATCACACCACA GTGAGGAGTGGTGGCAGGACTCGTGGGCCCTCTCCGGCCTGTTTCCCTGCCACTCCAGCCCC CTTGAGCTGGGAACTCAGGCTCCTGGAAAACTGGGCAGTGGGAGGCTCAGCTGCGGGCCAT TGATTGAGCCTTTGAGGGAGGATAGGGCTGGCCTTTGTGAAGCCAGCAGAGCTGAGAACCT CAGGCTTCCCTAGATCCAGAGCCCTCCCATCT		

	TCATCTCCAAGTTGGCCTGTGTCAGGCCATGAGGCAGCTGCACCCAGAGGCTATTGCAGCCATCCAGAGCAAGGCCCTGTTTAGCAAGGCTGAGGAGCAGCTGCTGAGCAAAGTGGGATCGACCAGCCAGCCACCTTGGAGACCTTCCAGGACCTGCTGCACAGACACCTGATGCCTTCTACCTGGCCGTACCGCGAAGGCCCTGCAGGCCCACTGGCAGCTCATGAAGCAGTATTACCTGCTGGAGGACAGACAGTGCAGCCGCTGCCCAAAGGGGACCAAGTGCTGAACTTCTCTGATGCAGAGGACCTGATTGATGACAGTAAGCTCAAGGACATGCGAGATGAGGTCTTGGAACTGAGCTGATGGTGGCTGACCGGCGCCAGAAGCGAGAGATTTCGGCAGCTGGAACAGGAAGTGCATAAGTGGCAGGTGCTAGTGGACAGCATCACAGGACATGAGCTCTCCGACTTCGACAACCAGACACTGGCAGTGTGCGGGGCCGCATGGTGGGTACCTGATGCGCTCGCGTGAGATCACCTGGGCAGAGCAACCAAGGATACCAGATTGATGTGGACCTGTCTCTGGAGGGTCCGGCCTGGAAGATATCCCGGAAACAAGGTGTCATCAAGCTGAAGAAACAACGGTGATTTCTTCATTGCCAATGAGGGTCGACGGCCCATCTACATCGATGGACGGCCGGTGCTCTGTGGCTCCAAATGGCGCCTCAGCAACAACCTCTGTGGTGGAGATCGCCAGCCTGCGATTCTGCTTCCCTTATCAACCAGGACCTCATTGCCCTCATCAGGGCTGAGGCTGCCAAGATCACACCACAGTGAGGAATGGTGGCAGGACTCGTGGGCCCTCTCCGGCCTGTTCCCCTGCCACTCCAGCCCCCTTGAGCTGGGAACTCAGGCTCCTGGAAAACTGGGCAGTGGGAGGCTCAGCTGCGGGCCATTGATTGAGCCTTTGAGGGAGGATAGGGCTGGCCTTTGTGAAGCCAGCAGAGGGCTGAGAACCTCAGGCTTCCCTAGATCCAGAGCCCCCTCCCATCTTCTCTCTCTCTAAAAACAACCTACCCCCATTCTACCCCCATTGCCACCTTCACTCCTGTGTCTCCAGCTGATTAGCCTCAGACTCTTCTTTATTGTTTTCTTTGTAAATAAAAAGCACCAGGTTCCAAAGTAAAAAAAAAAAAAAAAAATCGAG		
	ORF Start: ATG at 147		ORF Stop: TGA at 1533
	SEQ ID NO: 154	462 aa	MW at 51802.6kD
NOV37c, CG120401-03 Protein Sequence	MDKDSQGLLDSSLMASGTASRSEDEESLAGQKRASSQALGTIPKRRSSSRFIKRKKFDDELVESSLAKSSTRAKGASGVPEGRCSGSEPSSEKKKVSAPSTPVPPSPAPAPGLTKRVKKSQPLQVTKDLGRWKPADDLLLINAVLQTNDLTSVHLGVKFSRFTLREVQERWYALLYDPVISKLACQAMRQLHPEAIAAIQSKALFSAEEQLLSKVGSTSQPTLETQDILLHRHPDAFYLARTAKALQAHWQLMKQYLLLEDQTVQPLPKGDQVLNFSDAEDLIDDSKLKDMRDEVLEHELMVADRRQKREIRQLEQELHKWQVLVDSITGMSSPDFDNQTLAVLRGRMVRYLMRSREITLGRATKDNQIDVDLSLEGPWKISRQGVIKLKNNGDFFIANEGRRPIYIDGRPVLCGSKWRLSNNSVVEIASLRFVFLINQDLIALIRAEAAKITPO		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 37B.

Table 37B. Comparison of NOV37a against NOV37b and NOV37c.		
Protein Sequence	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV37b	77..534 5..419	371/458 (81%) 371/458 (81%)
NOV37c	77..534 5..462	415/458 (90%) 415/458 (90%)

5

Further analysis of the NOV37a protein yielded the following properties shown in Table 37C.

Table 37C. Protein Sequence Properties NOV37a	
PSort analysis:	0.9600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)

SignalP analysis:	No Known Signal Sequence Predicted
-------------------	------------------------------------

A search of the NOV37a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 37D.

5

Table 37D. Geneseq Results for NOV37a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAY77555	Human MIF1 protein (plasmid pCM577) - Homo sapiens, 462 aa. [WO200005362-A1, 03-FEB-2000]	77..534 5..462	458/458 (100%) 458/458 (100%)	0.0
AAY77554	Human MIF1 protein (plasmid pCM480) - Homo sapiens, 390 aa. [WO200005362-A1, 03-FEB-2000]	150..534 6..390	384/385 (99%) 385/385 (99%)	0.0
ABB57874	Drosophila melanogaster polypeptide SEQ ID NO 414 - Drosophila melanogaster, 578 aa. [WO200171042-A2, 27-SEP-2001]	160..530 204..572	209/375 (55%) 268/375 (70%)	e-111
AAY12243	Human 5' EST secreted protein SEQ ID NO: 556 - Homo sapiens, 42 aa. [WO9906554-A2, 11-FEB-1999]	77..113 5..41	37/37 (100%) 37/37 (100%)	1e-11
AAG40260	Arabidopsis thaliana protein fragment SEQ ID NO: 49928 - Arabidopsis thaliana, 465 aa. [EP1033405-A2, 06-SEP-2000]	406..521 328..443	36/116 (31%) 64/116 (55%)	9e-10

In a BLAST search of public sequence databases, the NOV37a protein was found to have homology to the proteins shown in the BLASTP data in Table 37E.

Table 37E. Public BLASTP Results for NOV37a				
Protein Accession Number	Protein/Organism/Length	NOV37a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75497	Cell cycle-regulated factor p78 - Homo sapiens (Human), 534 aa.	1..534 1..534	534/534 (100%) 534/534 (100%)	0.0
Q96EZ8	Unknown (protein for MGC:19577) - Homo sapiens (Human), 462 aa.	77..534 5..462	458/458 (100%) 458/458 (100%)	0.0

O14742	Nucleolar protein - Homo sapiens (Human), 462 aa.	77..534 5..462	457/458 (99%) 457/458 (99%)	0.0
Q99L90	Similar to microspherule protein 1 - Mus musculus (Mouse), 462 aa.	77..534 5..462	450/458 (98%) 451/458 (98%)	0.0
O35255	Nucleolar protein - Mus musculus (Mouse), 462 aa.	77..534 5..462	444/458 (96%) 450/458 (97%)	0.0

PFam analysis predicts that the NOV37a protein contains the domains shown in the Table 37F.

Table 37F. Domain Analysis of NOV37a			
Pfam Domain	NOV37a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FHA	435..508	16/82 (20%) 62/82 (76%)	6.2e-14

5

Example 38.

The NOV38 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 38A.

Table 38A. NOV38 Sequence Analysis			
	SEQ ID NO: 155	5142 bp	
NOV38a, CG122125-01 DNA Sequence	CCTGATGTGCAAGTACCAACATCTGTAAAAGATATGCGCTATTGCCAGGTTTCATTCCAAGAT GATCATGTGTCTTTGGAAAGTGCCTTTACAGTAAGACCACTTCCTGATGAACCTAAACATTTA AAATGTGAAATGAAAGGAGGAAAAACAGTACAGATGGGCCAAGAGCTTCAAGGAGAAGTAGTT ATAATAATTACAGATCAGTACGGAAATCAGATTCAAGCATTTTCACCAAGTTCTTTATCTTCT TTGTCAATTGCTGGGGTTGGACTTGATAGCTCAAATTTGAAAACAACCTTTCAGGAAAAACACA CAGAGTATAAGTGTAAGAGGCATCAAATTTATCCAGGTCCTCCTGGAAATAAGGATCTTTGT TTTACTTGGCGTGAGTTTCTGACTTTATTCGAGTGCAACTAATTTCTGGACCTCCTGCTAAA CTTCTCCTTATAGACTGGCCAGAATAAAGGAGTCCATTCCAGTGATTGGAAGAGATTTACAG AACCTATTATTGTTCAACTTTGTGATCAGTGGGATAATCCAGCACCGGTACAACATGTTAAA ATAAGTCTTACAAAAGCTAGCAATTTAAAGCTCATGCCTTCAAACCAACAGCATAAAACAGAT GAGAAAGGCAGGGCTAATTTGGGAGTATTCAGTGTTTTTGCCCTAGGGGAGAGCATACTCTT CAGGTTAAAGCCATCTATAACAAAAGTATCATAGAAGGACCTATAATTAAGTTAATGATTCTT CCAGACCCAGAAAAACCGGTTCTGCTCAATGTTAAATATGACAAAGATGCATCCTTCTTAGCA GGGGTCTTTTCACTGATTTTATGATTAGTGTATTCTGAAGATGACAGTATCATTAATAAAC ATTAATCCAGCACGTATTTCCATGAAATGTGGAAGCTGTCTACCAGTGGGAACCGACCCCA GCAATGCAGAAACATTTAGTTGTAATAAAATAAAAGATAATGACAAAGAAGATGGCTGCTTC TATTTAGGGATAAAGTAATTCCTAATAAAGTGGGGACATATTGTATCCAGTTTGGTTTTATG ATGGATAAAACAAATATTCTCAACAGTGAACAGGTTATAGTTGAAGTCTGCCTAATCAACCT GTGAAGTTAGTACCTAAAATTAAACCACCTACACCAGCTGTTTCAAATGTTTCGCTCAGTTGCC AGTAGGACCTTGGTCAGAGATCTACATCTTAGTATCACGGATGACTACGACAACCATACTGGA ATTGATTTGGTTGGCACTATAATAGCCACCATTAAAGGCTCTAATGAGGAAGATACTGATACC CCACTTTTTATTGGGAAAGTTAGAACACTTGAATTCCTTTCGTAATGGTTTCGGCTGAAATC ATGAGTCTGGTGTGCGAGAAAGTAGTCTGGAAGGATAGTACTGAATATTTTATTGTATT GAGCCCCGGCTACCACTTTTCAAGAACCTTAGAACCATATATCTACCGTTTCATGTTTTAC AATGATGTTAAGAAGCAGCAACAATGGCAGCACTTACAAAAGAAAAGGACCAATTATCTCAG TCTATTGTTATGTATAAAAGTTATTGGAAGCCAGCCAACAGCTTCTTAATGAAATGAAATGT		

	CAAGTTGAAGAAGCAAGATTAAAAGAGGCCCAATTGCGAAATGAACTAAAAATACATAATATT GACATTCCCTACAACACAACAGGTGCCACACATTGAAGCACTTCTGAAAAGAAAGCTATCAGAA CAAGAAGAACTGAAGAAAAACCTAGAAGATCGTGTACTCTTCCAACTATACTAAAGGCAGT GGAGATGTTTTGGGAAAGATTGCACATCTAGCACAAATTGAAGATGATAGAGCTGCGATGGTT ATTTCTTGGCATCTGGCAAGTGACATGGACTGTGTAGTACCCCTAACCACTGACGCTGCACGT CGTATCTATGATGAAACCCAAAGGTCGTGAGCAGGTGTGCCCCCTTGATTCTATTTACAAGAAG ACTCTTCCAGATTGGAAAAGATCTCTACCTCATTTCGGAATGGAAAATGTATTTTAAACCC ATTGGAGATCCAGTCTTTGCTCGAGACTTGTTAACATTTCCAGATAATGTAGAACATTGTGAA ACAGTATTTGGTATGCTGTGTAGGAGACACCATTTATTGGATAATCTGGATGCGGCCAATCAT TATAGAAAAGAGGTTGTTAAAATTACACACTGTCCTACACTGCTGACCAGAGATGGAGATCGA ATTGGAAGTAATGGAAAGTTTGGGGGCCTTCAGAATAAAGCTCTCCAATGGATAAACTTCGG GGAATGGTATTTGGAGCTCCAGTTCAAAAACAGTGTCTGATCTTAGGGGAACAAATAGATCT CTTCAGCAGTATCGTTCTGCTGTGTGCAAACTAGACAGTGTGAATAAGGATCTTAACAGTCAA TTAGAGTACCTTCGCACTCCGATATGAGGAAGAAAAAGCAAGAACTTGATGAACATGAGAAA AATCTCAAACTAATAGGAAAAAACTAGGTATGACTCCCATACGTAAGTGTATGACTCATTG CGTCATTACCAAAGGTTGAGACGACAGATTGTCCAGTTCCTCTAAAAGAATGAGACGAGAA GCTACAAGACAAAATAGGATTATAACCAAAACAGATGTATGAGAGGTGACAGAGAGAAGAGGC CATTGGTCTCAGTAAGAATGCCCTGCTTTCTGCATCTCTGTTTCAGAAGACCAAGAGGGTGAC TTACCAGACTGAGTATTTCTGGGACAATACAAGTACCTGGGCATGAATTTCCATTTTCGATT AGATGGGACTGGAAACAACCATTTCAATTTTATGAATCTTACTGGACATTATGGATTTACTGGA ATTATTCAGACATTATGCCCTTTGGTTGTCACTACCTTGCAAAATGTGTAAGAGGAAAAATGTG CTAATGTGGCAGTGACTGTAAAACCTGGCACATGGCATTATTATTAATCCTGAAGAAAAGTACATG TACTATTTTTCAGTATAAATATAATGAACATGTCAGAACTATTTCTTGAAAACCTTTTATTA CTTTTGGCGTAATTTTATTAACAAGATGTTTTGTCTTTTGTGTAAGGGAGGTTCTAGAGGCT AGATGTTTAATTGTAATATGTGAGGAAACTCAATGCAGAATTCAGGATAAAAAATTTTAAAG CACAGGTATTTGGGAATTGAAATGTTAAGATACCCAGAACAACATTAAATCAATGAGTGAAT TGTGACAGTGGTAGCATTTCAAATTTCAAAGACTTATCCTCTGTGTGTGTGTGTGTGTATAT ATATATATATATATAAATATATATATAAATATTCAGCAGCACCAAGTTTATACTATTG TTTGTTTGACTTTTATTAATACTAGAATATGTAGTCTCAGCCTTAATTTTACATTTACATTATT TTGTAATTTTTTATTACTATTTTAAAGGGTTAAAGAGAACATACATTCTCACATTAGTGTAC TTTCTGGTAGAAAGTTGCTGCAAAAACATTTGAAATGTATATTAACCTAATGTATGTATATA TATGTCCTTTGTGTAAGTTCAAGACTATTGATCTGTGAAGTTATTTTGTGAAGACATACATTTG GTAAGTAAGTTTGTGTGCCAGGAAATGTATGTGTTTTTAAACCTTTCTAAATATGCAGGCCA TTAATAAATAAGATTGTTTCTTCCCTACTGAATAGATAAGTGTTTTTCTTTTTTAAATTTGGA AGCTTCATAAAAGTTATCTTGTTAAAAAACGATGATGATGTTAACCTATCTTTATAATTGGAA ATTATTTAAACTGTTTGTGTTTACAGAAGAAACAAAATGGTAATTACAGAATTAGCTGTGGGG AAGATTGGCTCCCATGGTACTACAGGTTGAGTATCCCTTATTCAAATGGTTGGGACCTGAAG TGTTTTGGATTTTGGTTTTTGTGGGGTTTTTAAATTATTATTTTGGAATATTTGCATTATAT TTACCAATTTAGCATCCCTAATCAAAAATCTTTCATTAGAAATTGGAATGCTTCAATGAGC ATTTCCCTTATGTGTATCTGCGCTCAAATGTTTTGAATTTTGGAACTTTCTTGTTCTG ACTTTAGATTAGGGATACTACCCCGTACCAATATTGGAGTCTTTAGACATTAGATTATAT GAAAATGATTGATTGATAGGTAAGAAAGGTTAAAGAGAAATGACTTTTAGGAACATAGACTTG AACGTATAATTAATATGTGGAAGTGTGCTTTTATAATTCATTGTATCAAGAAGGAGTCAT GTTTTAAATCCAAGGCATTGAGACTATCTAGAAGCAAAAAGTTGGTTTTTAAAAATCTTTTT ATTAGAAGTGTGTTTTAAAAACATGTCTTTTTTCCCTTGAGCGACAGTGGCCTCACTACATTC CCCAGGCTAGTCTTGAACCTCCTGGGTTCAAGCAGTCTTCTGCTTGGCCTCCTAAGTAGCAG GGATTACAGGCATGCACCCACATGCCAATTTTTATTGCTTAAATGGCGAGTTCTATGGATAT TATCTAGCCCCAAGCCCATATCTGTACACTTTTGTTCAATTTTATAATACCAGAAAGAAATGTTT TATCAGTAAAAGAAGGTTTTAAGTATGAATCTCCATTTTTGTGGAGTTTTTCTACCTGTAAA AATTAATCTTTGGTTCCCGTAATTACAGTAAGGAGTTTTTCTATATTTTACAGTTGGATT TATCTAAGGATATTTCTCAGCATATTAAGGAATCCTTTTCTACATTTGAGCAAATACTGAGGT TCATGTTGTACCAAATAATAAATGCTTTTGTGTTTAAATATGTAACCGTAAGAACAATT GAAATTTTCTTCTAAGATTTAATACTAGTCTTTTAGTATAAAGGAATGTAAGTGGAGAAAAA ATTAATGCTACAGCCATTTATCCTGTGTTATGTGTTATATAATCTGAAATTTGTCAAAATGTT ACAAATCAGGAAGTTTTTTTTCTGGTTTCTTCTCCAAAGGAAAAATAAGCCTACTGTATT TGAAGTAACTGAAATAAACTCTCTCCAAGCCTTTTTGC		
	ORF Start: ATG at 34	ORF Stop: TGA at 2686	
	SEQ ID NO: 156	884 aa MW at 99946.2kD	
NOV38a, CG122125-01 Protein Sequence	MRYCQVSFQDDHVSLESFAFTVRPLPDEPKHLKCEMKGKTVQMGQELQGEVVIITDQYGNQI QAFSPSSLSSLSIAGVGLDSSNLKTTFQENTQISVRGIKFIPIPPGNKDLCTWREFSDFIR VQLISGPPAKLLLIDWPELKESIPVIGRDLQNPPIVQLCDQWDNPAPVQHVKISLTAKSNLKL MPSNQHKHTEKGRANLGVSVFAPRGEHTLQVKAIYNKSIIEGPIIKLMILPDPEKPVRLNV KYDKDASFLAGGLFTDFMISVISEDSSIKNINPARISMKMWLSTSGNRPPANAETFSCNKI KDNDKEDGCFYFRDKVIPNKVGTYCIQFGFMDKTNILNSEQVIVEVLNPQPVKLVPIKPP PAVSNVRSVASRTLVRDLHLSITDDYDNHTGIDLVTIIATIKGSNEEDTDTPLFIGKVRTLE		

PPFVNGSAEIMSLVLAESSPGRDSTEYFIVFEPRLPLLSRTLEPYILPFMFYNDVKKQQQMAA LTKEKDQLSQSIVMYKSLFEASQQLNEMKCQVEEARLKEAQLRNLKIHNDIPTTQQVPHI EALLKRKLSEQEELKKPRRSCTLPNYTKGSGDVLGKIAHLAQIEDDRAAMVISWHLASDMDC VVTLTDAARRIYDETQGRQQVLPDSIYKKTLPDWKRSLPHFRNGKLYFKPIGDPVFARDLL TFPDNVEHCETVFGMLLGDITILDNLDAANHYRKEVVKITHCPTLLTRDGDRIKSNKFGGLQ NKAPPMDKLRGMVFGAPVPKQCLILGEQIDLLQQYRSVCKLDSVNKDLNSQLEYLRTPDMRK KKQELDEHEKNLKLIEEKLGMTPIRKCNDLSLRHSPKVETTDPCVPPKMRREATRQNRITKT DV
--

Further analysis of the NOV38a protein yielded the following properties shown in Table 38B.

Table 38B. Protein Sequence Properties NOV38a	
PSort analysis:	0.9600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV38a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 38C.

Table 38C. Geneseq Results for NOV38a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG14416	Novel human diagnostic protein #14407 - Homo sapiens, 1482 aa. [WO200175067-A2, 11-OCT-2001]	1..884 577..1482	884/906 (97%) 884/906 (97%)	0.0
ABB97256	Novel human protein SEQ ID NO: 524 - Homo sapiens, 851 aa. [WO200222660-A2, 21-MAR-2002]	35..884 1..851	850/851 (99%) 850/851 (99%)	0.0
ABG20071	Novel human diagnostic protein #20062 - Homo sapiens, 848 aa. [WO200175067-A2, 11-OCT-2001]	38..884 1..848	847/848 (99%) 847/848 (99%)	0.0
AAE01527	Human gene 15 encoded secreted protein fragment, SEQ ID NO:184 - Homo sapiens, 646 aa. [WO200134626-A1, 17-MAY-2001]	239..884 1..646	646/646 (100%) 646/646 (100%)	0.0
ABG20070	Novel human diagnostic protein #20061 - Homo sapiens, 1300 aa. [WO200175067-A2, 11-OCT-2001]	123..785 273..830	459/663 (69%) 484/663 (72%)	0.0

In a BLAST search of public sequence databases, the NOV38a protein was found to have homology to the proteins shown in the BLASTP data in Table 38D.

Table 38D. Public BLASTP Results for NOV38a				
Protein Accession Number	Protein/Organism/Length	NOV38a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O75141	K1AA0650 protein - Homo sapiens (Human), 848 aa (fragment).	38..884 1..848	847/848 (99%) 847/848 (99%)	0.0
Q9D4M7	4931400A14Rik protein - Mus musculus (Mouse), 887 aa.	2..877 3..879	719/877 (81%) 809/877 (91%)	0.0
Q9UG39	Hypothetical 82.6 kDa protein - Homo sapiens (Human), 728 aa (fragment).	157..884 1..728	727/728 (99%) 728/728 (99%)	0.0
Q9H6Q2	CDNA: FLJ21993 fis, clone HEP06576 - Homo sapiens (Human), 267 aa.	618..884 1..267	267/267 (100%) 267/267 (100%)	e-157
Q8T1V3	Hypothetical 255.5 kDa protein - Dictyostelium discoideum (Slime mold), 2284 aa.	246..644 1646..2070	111/455 (24%) 182/455 (39%)	2e-08

5

PFam analysis predicts that the NOV38a protein contains the domains shown in the Table 38E.

Table 38E. Domain Analysis of NOV38a			
Pfam Domain	NOV38a Match Region	Identities/ Similarities for the Matched Region	Expect Value

10

Example 39.

The NOV39 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 39A.

Table 39A. NOV39 Sequence Analysis			
	SEQ ID NO: 157	9195 bp	
NOV39a, CG122195-01 DNA	ATGGCCAAGTATGGAGAACATGAAGCCAGTCCTGACAATGGGCAGAACGAATTCAGTGATATC ATTAAGTCCAGATCTGATGAACACAATGACGTACAGAAGAAAACCTTTACCAAATGGATAAAT		

Sequence	GCTCGATTTTCAAAGAGTGGGAAACCACCCATCAATGATATGTTACAGACCTCAAAGATGGA AGGAAGCTATTGGATCTTCTAGAAGGCCTCACAGGAACATCACTGCCAAAGGAACGTGGTTCC ACAAGGGTACATGCCTTAATAACGTCAACAGAGTGTCTGCAGGTTTTACATCAGAACAATGTG GAATTAGTGAATATAGGGGAACTGACATTGTGGATGGAAATCACAACTGACTTTGGGGTTA CTTTGGAGCATCATTTTGCAGTGGCAGGTGAAAGATGTCATGAAGGATGTCATGTCGGACCTG CAGCAGACGAACAGTGAGAAGATCTCTGCTCAGCTGGGTGCGTCAGACCACCGAGCCCTACAGC CAAGTCAACGTCTCAACTTCAACCACAGCTGGACAGATGGACTCGCCTTTAATGCTGTCTC CACCGACATAAACCTGATCTCTCAGCTGGGATAAAGTTGTCAAATGTCACCAATTGAGAGA CTTGAACATGCCTTCAGCAAGGCTCAAACCTATTTGGGAATTGAAAAGCTGTTAGATCCTGAA GATGTTGCCGTTCCGGCTTCTGACAAGAAATCCATAATTATGTATTTAACATCTTTGTTTGTGAG GTGCTACCTCAGCAAGTCACCATAGACGCCATCCGTGAGGTAGAGACACTCCCAAGGAAATAT AAAAAAGAAATGTGAAGAAGAGGCAATTAATATACAGAGTACAGCGCTGAGGAGGAGCATGAG AGTCCCGAGCTGAACTCCAGCACTGTCACTGAGGTGACATGGATCTGGACAGCTATCAG ATTGCGTTGGAGGAAGTGTGACCTGGTTGCTTTCTGCTGAGGACACTTTCCAGGAGCAGGAT GATATTTCTGATGATGTTGAAGAAGTCAAAGACCAGTTTGCAACCCATGAAGCTTTTATGATG GAACTGACTGCACACCAGAGCAGTGTGGGCAGCGTCTGCAGGCAGGCAACCAACTGATAACA CAAGGAACCTCTGTGACAGGAAGAAGAAATTGAGATTCAGGAACAGATGACCTGTGTAATGTG AGATGGGAGGCTCTTAGGGTGGAGAGTATGGACAGACAGTCCCGGCTGCACGATGTGCTGATG GAACTGCAGAAGAAGCAACTGCAGCAGCTCTCCGCTGGTTAACTCAGAGAGGAGCGCATT CAGAAGATGGAACTTGCCCTGGATGATGATGTAAATCTCTACAAAAGCTGCTAGAAGAA CATAAAGTTTGCAAAGTGATCTTGAGGCTGAACAGGTGAAAGTAAATTCATAACTCACATG GTGGTCATTGTTGATGAAAACAGTGGTGAGAGCGCTACAGCTATCCTAGAAGACCATGTACAG AACTTGGTGAGCGCTGGACAGCAGTATGCCGTTGGAAGTGAAGAACGCTGGAAATAGGTTACAA GAAATCAATATATTGTGCGAGGAATTATTGGAAGAAGTGTGTTGAAAGCTTGGTTAACC GAAAAAGAGAGGCTTTAAATAAAGTCCAGACAAGCAACTTCAAAGACCAAGGAAGCACTAAGT GTCAGTGTTCGACGTCTGGCTATTTGAAGGAAGACATGGAAATGAAGCGTCAAACATTGGAT CAGCTGAGTGAGATTGGCCAGGATGTGGGACAATTACTTGATAATTCCAAGGCATCTAAGAAG ATCAACAGTGACTCAGAGGAAGTCACTCAAAGATGGGATCTTTGGTTTCAAGAGACTAGAAGAT TCCTCCAACAGGTGACTCAGGCTGTAGCAAAGCTGGGGATGTCTCAGATTCTCAGAAGGAC CTTTTGGAGACTGTTCTGTGAAGAGAACAAGCAATTACAAAAAATCTAAGCAGGAAGTGCCT CTCCTCTCTCCCAAGAGAGACAGATCCATGTGGATATGAAGCTAAGAAAGGCTTTGAT GCTATAAGTGACAGAGCTGTTGAAGTGGATTTTGAATGGAAAAGTCCATCAGACCAAGAG ATAAAGAGTATATGAAGATGCAAGACACTTCCGAAATGAAAAGAGTTGAAGGCATTAGAA AAAGAACAGAGAGAAAGAAATCCCAAGAGCAGATGAATTAACCAAGTGGACAAATCCTTGTG GAGCAATGGGAAAAGAGGCTTCTACTGAAGAAATAAAAAATGTTCTGGAGAAGGTTTCA TCAGAATGGAAGAATGTATCTCAACATTTGGAAGATCTAGAAAAGAAAGATTGAGTACAGGAA GATATAAATGCTTATTTCAAGCAGCTTGATGAGCTTGAAGAGTCAAGACAAAGGAGGAG TGGGTAAAACACACTTCCATTTCTGAATCTTCCCGGCAGTCTTGCCAAGCTTGAAGGATTCC TGTCAGCGGGAATTGACAAATCTTCTGGCCTTCAACCAAAATGAAATGGCTCGTGCAAGC TGCTCGGCCCTGATGTCTCAGCCTTCTGCCAGATTGTTGTCAGCGGGCTTCGATAGCTTT CTGGGCCGCTACCAAGCTGTACAAGAGGCTGTAGAGGATCGTCAACAACATCTAGAGAATGAA CTGAAGGGCCAACTGGACATGCATATCTGGAACATTGAAAACACTGAAAGATGTGCTAAAT GATTGAGAAAATAAGGCCAGGTGTCTCTGAATGTCCTTAATGATCTTGCCAAGGTGGAGAAAG GCCCTGCAAGAAAAAGACCTTGTGAAATCCTTGAGAATCAGAACTGCATTACATAAA CTTGCAAGAAACAAAGGCTCTGGAGAAAAATGTTTATCTGATGTAGAAAAATATATAAG CAAGAATTTGATGATGTGCAAGGAAAGTGAACAAGCTAAAGGTCTTGGTTTCAAAGATCTA CATTTGCTTGAGGAAATGTCTCTCACTCAGAGCTTTTGAGGCCGATTCAACAGTCATTGAG AAGTGGATGGATGGCGTGAAAGACTTCTTAATGAAACAGCAGGCTGCCAAGGAGACGACGA GGTCTACAGAGGCAGTTAGACCAGTGCTCTGCATTTGTTAATGAAATAGAAACAATTGAATCA TCTCTGAAAAACATGAAGGAAATAGAGACTAATCTTGAAGTGGTCCAGTTGCTGGAAATAAAA ACTTGGGTGCAGACAAGACTAGGTGACTACCAACTCACTGGAGAACTTAGCAAGGAGATC GCTACTCAAAAAGTAGGTTGTCTGAAAGTCAAGAAAAAGCTGCGAACCTGAAGAAAGACTTG GCAGAGATGCAGGAATGGATGACCCAGGCGAGGAAGAATATTGGAGCGGGATTTGAGTAC AAGTCACAGAGAGCTTGAGAGTGCTGTGGAAGAGATGAAGAGGGCAAGAGAGGATGTGTTG CAGAAGGAGGTGAGAGTGAAGATTCTCAAGGACAACATCAAGTTATTAGTGCCTAAGGTGCC TCTGGTGGCCAGGAGTTGACGTCTGAGCTGAATGTTGTGCTGGAGAATTACCAACTTCTTTGT AATAGAATTCGAGGAAAGTGCCACACGCTAGAGGAGGTCTGGTCTTGTGGATTGAACGTCTT CACTATTGGATCTTGAAGTCACTGGTTAAACACTTTGGAAGAGCGGATGAAGAGCACAGAG GTCCTGCCTGAGAAGACGGATGCTGTCAACGAAGCCCTGGAGTCTCTGGAATCTGTTCTGCCG CACCCGCGCAGATAATCGACCCAGATTCTGAGAGCTTGGCCAGACTCTGATTGATGGGGGATC CTGGATGATATAATCAGTGAGAACTGGAGGCTTTCAACAGCCGATATGAAGATCTAAGTAC CTGGCAGAGAGCAAGCAGATTCTTTGGAAGCAACTCCAGGTGCTGCGGAACTGCACGAG ATGCTTCAAGTCTTGCAAGAGAGCTTGGGGAGCTGGACAAACAGCTCACCACATCTGACT GACAGGATAGATGCTTTCAAGTCCACAGGAAGCTCAGAAAATCCAAGCAGAGATCTCAGCC CATGAGCTAACCTAGAGGAGTTGAGAAGAAATATGCGTTCTCAGCCCTGACCTCCCAAGAG AGTAGGACTGCCAGAGGAGGAAGTCAGATGGATGTGCTACAGAGGAACTCCAGAGGTTGTC
----------	--

ACAAAGTTCCAGCTTTTCCAGAAGCCAGCTAACTTCGAGCAGCGCATGCTGGACTGCAAGCGT
GTGCTGGATGGCGTGAAAGCAGAACTTCACGTCTGGATGTGAAGGACGTAGACCTTGACGTC
ATACAGACGCACCTGGACAAGTGTATGAACTGTATAAACTTTGAGTGAAGTCAAACTTGAA
GTGGAACTGTGATTAAACAGGAAGACATATTGTCCAGAAACAGCAACGGGACAAACAAAA
GGGATGGATGAGCAGCTGACTTCCCTGAAGGTTCTTTACAATGACCTGGCGCACAGGTGACA
GAAGGAAACAGGATCTGGAAGAGCATCACAGTTGGCCCGGAAAATGAAGAAAGAGGCTGCT
TCTCTCTCTGAATGGCTTTCTGCTACTGAACTGAATTGGTACAGAACTCCACTTCAGAAAGT
CTGCTTGGTGACTTGGATACAGAAATTTCTGGGCTAAAAATGTTCTGAAGGATCTGGAAGG
AGAAAAGCTGATTTAAATACCATCACAGAGAGTAGTGCTGCCCTGCAAACTTGATTGAGGGC
AGTGAGCCTATTTAGAAGAGAGGCTCTGCGTCTTAACGCTGGGTGGAGCCGAGTTCTGTAAC
TGGACTGAAGATTGGTGCAATACCTTGATGAACCATCAGAACCAGCTAGAAATATTGTAGGG
AACGTGGCTCACATAAGTACCTGGCTTTATCAAGCTGAAGCTCTATTGGATGAAATTGAAAA
AAACCAACAAGTAACAGGAAGAAATGTGAAGCGTTTAGTATCTGAGCTGGATGATGCCAAC
CTCCAGGTTGAAATGTCCCGCATCAAGCCCTTATTTGTATGAATGCCGTGGAAGCTCAAGC
AGGGAGCTTGTAGAACCAAGTTAGCTGAGCTGAATAGGAACCTTGAAAAGGTGTCTCAACAT
ATCAAAAGTGCCAAATTGTAAATTGCTCAGGAACCATATACCAATGTTTGGTCACCACTGAA
ACATTTGAACTGGTGTGCTTTCTCTGACTTGGAAAAATTAGAAAAAGACATAGAAAAATG
TTAAAAATTTGTGGAAAAACACTTGAATCCAGTGATGAAGATGAAAAGATGGATGAGGAGAT
GCCCAGATTGAGGAAGTTCTACAAAGAGGAGAAGAAATGTTACATCAACCTATGGAAGATAAT
AAAAAGAAAAGATCCGTTTGCAATTATTACTTTTGCACTAGATACAACAAATTAAGGCA
ATCCCTATTCAACAGAGGAAAATGGGTCAACTTGCTTCTGGAATTAGATCATCACTTTCTCCT
ACAGATTATCTGGTTGAAATTAACAAAATTTACTTTGCTATGGATGATGTTGAATTATCGCTT
AATGTTCCAGAGCTCAACACTGCTATTACGAAGACTTCTCTTTTCAGGAAGACTCTCTGAAG
AATATCAAGACCAACTGGACAACTTGGAGAGCAGATTGCAGTCATTATGAAAAACAGCCA
GATGTCTATCTTGAAGCCTCTGGACCTGAAGCCATTAGATCAGAGATACACTTACTCAGCAT
GGCGTTGAGCTAAGACAGCAGCAGCTTGGAGCATGATTATTGACAGTCTTCAGTGGGATGAC
CATAGGGAGGAGACTGAAGAACTGATGAGAAAAATATGAGGCTCGACTCTATATTTCTCAGCAA
GCCCAGCGGATCCACTACCAACAAATTTCTGATAACCAAACTGCTTCAAGAAGTGGGT
CCTGGAGATGGTATCGTCTATGGCGTTCGGATACGTCTGCGAGAACTCTGGAGGGAATATGGG
AGTGATGACACAAGGAATGTGAAGAAACCAAGAGTACTTAAAAACATCATGGATCAATCTC
AAACAAAGTATTGCTGACAGACAGAACGCTTGGAGGCTGAGTGGAGGACGGTGCAGGCCCTC
CGCAGAGATCTGGAAGCTTCTGAAAGTGGATCCAAGAAGCAGAGACCAAGTGAATGTGCTT
GTGGATGCCTCTCATCGGAGAAATGCTCTTCAGGATAGTATCTTGGCCAGGGAACCTCAACAG
CAGATGCAGGACATCCAGGCAGAAATGATGCCCAATGACATATTAAAAAGCATTGACGGA
AACAGGCAGAAGATGGTAAAAGCTTTGGGAAATTTGAAGAGGCTACTATGCTTCAACATCGA
CTGGATGATATGAACCAAGATGGAATGACTTAAAGCAAAATCTGCTAGCATCAGGCCCTTGAAG
TTGGAGGCCAGCGCTGAGAAGTGGAAACAGGTTGCTGATGTCTTAGAAGAACTGATCAAAATGG
CTGAATATGAAAGATGAAGAGCTTAAAGAAACAAATGCCTATTGGAGGAGATGTTCCAGCCTTA
CAGCTCCAGTATGACCATTGTAAGGCCCTGAGACGGGAGTTAAAGGAGAAAGAAATTTCTGTCT
CTGAATGCTGTGACAGGCCCGAGTTTTCTTGGCTGATCAGCCAAATGAGGCCCTTGAAGAG
CCAAGAAGAAACCTACAATCAAAAACAGAAATTAACCTCTGAGGAGAGAGCCCAAGATTGCC
AAAGCCATGCGCAACAGTCTTCTGAAGTCAAAGAAAAATGGGAAAGTCTAAATGTGTAACT
AGCAATTGGCAAAAGCAAGTGGACAAGGCATTGGAGAACTCAGAGACCTGCAGGGAGCTATG
GATGACCTGGACGCTGACATGAAGGAGGCAGAGTCCGTGCGGAATGGCTGGAAGCCCGTGGGA
GACTTACTCATTGACTCGCTGACGATCAGATGAAAAAATCATGCAATTATGAGAAGAAAT
GCACCAATCAACTTTAAAGTTAAAACGGTGAATGATTTATCCAGTCAGCTGTCTCCACTTGAC
CTGCATCCCTCTCTAAAGATGTCTCGCCAGCTAGATGACCTTAATATGCGATGGAAGCTTTTA
CAGGTTTCTGTGGATGATCGCTTAAACAGCTTCAGGAAGCCACAGAGATTTTGGACCATCC
TCTCAGCATTTTCTCTACGTGAGTCCAGCTGCGGTGGCAAGATCCATTTCACATAATAAAA
GTGCCCTATTACATCAACCATCAAAACAGACCACCTGTTGGGACCATCTAAAATGACCGAA
CTCTTTCAATCCCTTGCTGACCTGAATAATGTACGTTTTTCTGCCTACCGTACAGCAATCAAA
ATCCGAAGACTACAAAAGCACTATGTTTGGATCTCTTAGAGTTGAGTACAACAAATGAAAT
TTCAAACAGCAAGTTGAACCAAAATGACAGCTCCTCAGTGTTCCAGATGTCATCAACTGT
CTGACAACAACTTATGATGGACTTGAGCAAAATGCATAAGGACCTGGTCAACGTTCCACTCTGT
GTTGATATGTGTCTCAATTGGTTGCTCAATGTCTATGACACGGGTGCAACTGGAAAAATTAGA
GTGCAGAGTCTGAAGATTGGATTAATGTCTCTCTCAAAGGTCTCTTGGAGAAAAATACAGA
TATCTCTTTAAGGAAGTTGCGGGCCGACAGAAATGTGTGACAGAGGACGCTGGGCTGTTA
CTTCATGATGCCATCCAGATCCCCCGGAGCTAGGTGAAGTAGCAGCTTTTGGAGGCAGTAAT
ATTGAGCCTAGTGTTCCGAGCTGCTTCCAACAGAAATAACAATAAACAGAAATAAGTGTGAAA
GAGTTTATAGATTGGATGATTTGGAACACAGTCCATGGTTTGGCTCCAGTTTTACATCGA
GTGGCAGCAGCGGAGACTGCAAAACATCAGGCCAAATGCAACATCTGTAAGAATGTCCAATT
GTCCGGTTCAGGTATAGAAGCCTTAAGCATTTTAACTATGATGTCTGCCAGAGTGTCTCTTT
TCGGGTGCAACAGCAAAAGGTCACAAATTACATTACCCAATGGTGAATATTGTATACCTACA
ACATCTGGGGAAGATGTACGAGACTTCAAAAGGTACTTAAAGAACAGTTGAGTGAAGAAG
TACTTTGCCAAACACCCTCGACTTGGTTACCTGCTGTCCAGACAGTTCTTGAAGGTGACAACT
TTAGAGACTCCTATCACACTCATCAGTATGTGGCCAGAGCACTATGACCCCTCACAACTCCTCT

	CAACTGTTTCATGATGACACCCATTCAAGAATAGAACAATATGCCACACGACTGGCCAGATG GAAAGGACTAATGGGTCTTTTCTCACTGATAGCAGCTCCACCACAGGAAGTGTGGAAAGACG CACGCCCTCATCCAGCAGTATTGCCAAACACTCGGAGGAGAGTCCCCAGTGAGCCAGCCGAC AGCCCAGCTCAGATCCTGAAGTCAGTAGAGAGGGAAGAAGCTGGAGAACTGGAGAGGATCATT GCTGACCTGGAGGAAGAACAAGAAATCTACAGGTGGAGTATGAGCAGCTGAAGGACGAGCAC CTCCGAAGGGGGCTCCCTGTCGGTTCCACGCCAGAGTCGATTATATCTCCCCATCACACGTCT GAGGATTCAAACTTATAGCAGAAGCAAACTCCTCAGGCAGCACAAAGGTCGGCTGGAGGCT AGGATGCAGATTTTAGAAGATCACAATAAACAGCTGGAGTCTCAGCTCCACCGCCTCCGACAG CTGCTGGAGCAGCCTGAATCTGATTCCCGAATCAATGGTGTTCCTCCATGGGCTTCTCCTCAC CATTCTGCAGCTGAGTACTCGCTTGATCCAGATGCCTCCGGCCACAGTTCACCAGGACGAGC GGAGAGGACCTGCTGGCCCCACCGCAGCAGACCAGCAGGATCTCACGGAGGTATGGAGCAG ATTACAGCAGCGTTTCCATCTTGCTGCCCAAATGTTCCAGCAGGCCACAGGCAATGTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 9193
	SEQ ID NO: 158	3064 aa	MW at 352283.8kD
NOV39a, CG122195-01 Protein Sequence	MAKYGEHEASPDNGQNEFSDIIKRSRDEHNDVQKKFTFKWINARFSKSGKPPINDMFTDLKDG RKLLDLLEGLTGTSLPKERGSTRVHALNNVNRVLQVLHQNNVELVNIIGTDIVDGNHKLTLGL LWSIILHWQVKDVMKDVMSDLQQTNSEKILLSWVRQTRPYSQVNVNFTTSTWDGLAFNAVL HRHKPDLFSWDKVVKMSPIERLEHAFSKAQTYLGLIEKLLDPEDVAVRLPKDKSIIMYLTSLFE VLPQQVTIDAIREVELTPRKYYKECEEEA INIQSTAPEEEHESPRAEPTSTVTEVDMDLDSYQ IALEEVLTWLLSAEDTFQEODDISDDVEEVKQDFATHEAFMMELTAHQSSVGSVLQAGNQLIT QGTLSDEEEFEIQEQMTLLNARWEALRVESMDRQSRLLHDVLMELQKKQLQQLSAWLTLEERI QKMETCPLDDDVKSLLQKLLLEHKSLSQSDLEAEQVKNVSLTHMVIVDENSSESATAILEDQLQ KLGERWTAVCRWTEERWNRLOEINILWQELLEEQCLLKAWLTEKEEALNKVQTSNFKDQKELS VSVRRLLAILKEDMEMKRQTLQDLSEIGQDVQGLLDNSKASKKINSDELTQRWDSLVRQLED SSNQVTQAVAKLGMSQIPQKDLLETVRVREQAITKSKQELPPPPPPKKRQIHVDIEAKKKFD AISAEELNWLKWKTAIQTTEIKEYMKMQDTSEMKKLKALEKEQRRERIPRADELNQTGQILV EQMGKEGLPTEEIKNVLEKVSSEWKNVSQLHEDLERKIQLOEDINAYFKQLEDEKVIKTKEE WVKHTSISESSRQSLPSLKDSCRELTNLLGLHPKIEMARASCSCALMSQSPAPDFVQRGFDSF LGRYQAVQEAVEDRQQHLENELKGQPGHAYLETTLKTLKDVLNDSENKAQVSLNVLNDLAKVEK ALQEKKTLDIEILENQKPALHKLAEETKALEKNVHPDVEKLYKQEFDDVQGGKNNKLKVLVSKDL HLEEIALTLRAFEADSTVIEKWMGDVKGDFLMKQQAQGGDAGLQRLDQCSAFVNEIETIES SLKNMKEIETNLRSGPVAGIKTWQTRLGDYQTLQLEKLSKEIATQKSRLESSEKAEANLKKDL AEMQEWMTQAEEEYLERDFEYKSPEELESASEEMKRAKEDVLQKEVRVKILKDNIKLLAAKVP SGGQELTSELNVVLENYQLLCNRIRGKCHTLEEVWSCWIELLHYLDLETTWNTLEERMKSTE VLPEKTDVNEALESLESVLRHPADNRTQIRELGQTLIDGGILDDIISEKLEAFNSRYEDLSH LAESKQISLEKQLQVLRQDQMLQVLQESLGELDKQLTTLTDRIDAFQVPQEAQKIQAEISA HELTLEELRRNMRSPQLTSPESRTARGGSQMDVLQRLKREVSTKFLQFQKPNFQRMMDCKR VLDGVKAELHVLVDKVDVDPVQIOTHLDKCMKLYKTLSEVKLEVETVIKTGRHIVQKQQTDPNK GMDEQLTSLKVLVYNDLGAQVTEGQDLERASQLARKMKKEAASLSEWLSATETELVQKSTSEG LLGDLDEISWAKNVLDLEKRAKADLNTITESSAALQNLIEGSEPILEERLCVLNAGWSRVRT WTEPDWNTLMNHQNLQLEIFDGNVAHISTWLYQAEALLDEIEKKPTSKQEEIVKRLVSELLDAN LQVENVRDQALILMNARGSSSRELVEPKLAELNRNFEKVSQHIKSAKLLIAQEPLYQCLVTTE TFETGVPSDLEKLENDIENMLKFVEKHLESSDEDEKMDSESAQIEEVLQRGEEMLHQPMEDN KKEKIRLQLLLLHTRYNKIKAIPIQQRKMQLASGIRSSLLPTDYLVEINKILLCMDDVELSL NVPELNTAIYEDFSQEDSLKNIKQDLKLEQIAVIEHKQPDVILEASGPFAIQIRDTLTQH GVELRQQQLEDMIIDSLQWDDHREETEELMRKYEARLYILQARRDPLTKQISDNQILLQELG PGDGVMAFGYVLQKLWREYGSDDTRNVKETTEYLKTSWINLKQSIADRQNALEAEWRTVQAS RRDLENFLKWIQEAETTVNVLVDASHRENALQDSILARELKQMQMDIQAEIDAHNDFKSIDG NRQKMVKALGNSEETMLQHRLLDMNQNRWDLKAKSASIRAHLEASAEKNRLLMSLEELIKW LNMKDEELKKQMPIGGDVPAQLQYDHCALRRELKEKEYSVLNAVDAQRVFLADQPIEAPEE PRRNLQSKTELTPERAQKIAKAMRKQSSEVKEKWESENAVTSNWQKQVDKALEKLRDLQGM DDLADAMKEAESVRNGWKPVGDLIDSLQDHIEKIMAFREEIAPINFVKVTNVDLSQSLPLD LHPSLKMSRQLDDLNRWKLLQVSVDDRKLQQLQEAHRDFGPSSQHFLSTSVQLPWQRSISHNK VPYYINHQTQTCWDHPKMTLQSLADLNNVRFSAYRTAIKIRRLQKALCLDLLELSTTNEI FKQHKLNQNDQLLSVPDVINCLTTTYDGLEQMHKDLVNVPLCVDMLNWLNNVYDTGRTGKIR VQSLKIGLMSLSKGLLEEKYRYLFKEVAGFTEMCDQRQLGLLLHDAIQIPRQLGEVAAFSSGN IEPSVRSCFQQNNKPEISVKEFIDWMHLEPQSMVWLPVLHRAAAETAKHQAKCNICKECPI VGFYRSLKHFNYDVCQSCFFSGRTAKGHKLHYPMVEYCIPTTSGEDVDRFTKVLKNKFRSKK YFAKHPRLLGYLPVQTVLEGDNLETPITLISMWPEHYDPSQSPQLFDDHTLSRIEQYATRLAQM ERTNGSFLTDSSTTGSVEDEHALIQYQCQLTGGESPVSPQSPAQILKSVEREERGERLII ADLEEEQRNLQVEYELKQDQHLRRLPVGSPPEIISPHHTSEDSSELIAEAKLLRQHKGRLEA RMQILEDHNNKQLESQHLRLQLEQPESDSRINGVSPWASPQHSALSYSLLDPDASGPQFHQAA GEDLLAPPHDTSTDLETEVMEQIHSTFPSCCPNVPSRPQAM		

Further analysis of the NOV39a protein yielded the following properties shown in Table 39B.

Table 39B. Protein Sequence Properties NOV39a	
PSort analysis:	0.9600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV39a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 39C.

Table 39C. Geneseq Results for NOV39a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW22017	Utrophin - Homo sapiens, 3433 aa. [WO9722696-A1, 26-JUN-1997]	1..2429 1..2462	1995/2509 (79%) 2115/2509 (83%)	0.0
AAB67964	Amino acid sequence of utrophin B isoform minigene - Unidentified, 2013 aa. [WO200125461-A1, 12-APR-2001]	1492..3064 371..2013	1215/1646 (73%) 1322/1646 (79%)	0.0
AAW22016	Utrophin truncated polypeptide - Synthetic, 2008 aa. [WO9722696-A1, 26-JUN-1997]	1492..3064 366..2008	1214/1646 (73%) 1322/1646 (79%)	0.0
AAP90373	Sequence encoded by human muscular dystrophy (MD) cDNA - Homo sapiens, 3685 aa. [WO8906286-A, 13-JUL-1989]	28..2435 12..2429	974/2498 (38%) 1497/2498 (58%)	0.0
AAP90290	Human Duchenne muscular dystrophy gene - Homo sapiens, 3685 aa. [EP331514-A, 06-SEP-1989]	28..2435 12..2429	962/2498 (38%) 1485/2498 (58%)	0.0

- 10 In a BLAST search of public sequence databases, the NOV39a protein was found to have homology to the proteins shown in the BLASTP data in Table 39D.

Table 39D. Public BLASTP Results for NOV39a				
Protein Accession Number	Protein/Organism/Length	NOV39a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P46939	Utrophin (Dystrophin-related protein 1) (DRP1) (DRP) - Homo sapiens (Human), 3433 aa.	1..2429 1..2462	2055/2509 (81%) 2152/2509 (84%)	0.0
CAA03735	SEQUENCE 9 FROM PATENT WO9722696 - unidentified, 3433 aa.	1..2429 1..2462	1996/2509 (79%) 2116/2509 (83%)	0.0
O08614	Cytoskeletal protein - Mus musculus (Mouse), 3429 aa.	1..2556 1..2540	1795/2616 (68%) 2028/2616 (76%)	0.0
O55147	Utrophin - Rattus norvegicus (Rat), 3419 aa.	1..2243 1..2266	1728/2297 (75%) 1911/2297 (82%)	0.0
CAC37761	Sequence 8 from Patent WO0125461 - synthetic construct, 2013 aa.	1492..3064 371..2013	1215/1646 (73%) 1322/1646 (79%)	0.0

PFam analysis predicts that the NOV39a protein contains the domains shown in the Table 39E.

Table 39E. Domain Analysis of NOV39a			
Pfam Domain	NOV39a Match Region	Identities/ Similarities for the Matched Region	Expect Value
CH	31..135	45/125 (36%) 91/125 (73%)	6.1e-35
CH	150..255	32/125 (26%) 92/125 (74%)	1.9e-35
spectrin	309..417	32/110 (29%) 77/110 (70%)	3.3e-17
spectrin	418..526	26/111 (23%) 82/111 (74%)	8.1e-14
spectrin	541..637	20/99 (20%) 65/99 (66%)	0.042
spectrin	687..798	28/113 (25%) 72/113 (64%)	0.67
spectrin	803..902	18/104 (17%) 68/104 (65%)	0.0039

spectrin	1016..1083	16/69 (23%) 45/69 (65%)	0.00029
spectrin	1125..1230	19/109 (17%) 73/109 (67%)	2.3e-10
DUF164	1069..1278	40/243 (16%) 111/243 (46%)	0.85
spectrin	1248..1334	29/93 (31%) 52/93 (56%)	0.0006
spectrin	1544..1649	18/109 (17%) 72/109 (66%)	1.4e-05
spectrin	1652..1753	20/108 (19%) 69/108 (64%)	0.065
spectrin	2037..2071	12/37 (32%) 27/37 (73%)	0.0066
spectrin	2074..2187	27/115 (23%) 82/115 (71%)	1.8e-09
spectrin	2190..2267	15/81 (19%) 56/81 (69%)	0.008
spectrin	2322..2428	27/110 (25%) 71/110 (65%)	4.8e-05
WW	2445..2474	10/30 (33%) 24/30 (80%)	5.1e-08
ZZ	2695..2740	21/47 (45%) 45/47 (96%)	7.5e-23

Example 40.

The NOV40 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 40A.

5

Table 40A. NOV40 Sequence Analysis			
	SEQ ID NO: 159	1677 bp	
NOV40a, CG122738-01 DNA Sequence	GCGCGACCGGCAGCCATGAGCTCGGAGATGGAGCCGCTGCTCCTGGCCTGGAGCTATTTTAGG CGCAGGAAGTTCCAGCTCTGCGCCGATCTATGCACGCAGATGCTGGAGAAGTCCCTTATGAC CAGGCAGCTTGGATCTTAAAGCAAGAGCGCTAACAGAAATGGTATACATAGATGAAATTGAT GTAGATCAGGAAGGAATTGCAGAAATGATGCTGGATGAAAATGCTATAGCTCAAGTTCACAGT CCTGGAACGTCTTTGAACTCCCTGGAATAATCAGACAGGAGGGCCTAGCCAGGCGGTTAGG CCAATCACACAAGCTGGAAGACCCATTACAGGTTTCCTCAGGCCACGACGCAGAGTGAAGG CCAGGCACTATGGAACAGGCTATCAGAACACCCAGAACCCTACACAGCCCGCCCTATCACC AGCTCCTCCGGAAGATTTGTGAGGCTGGGAACGGCTTCCATGCTTACAAGTCTGATGGACCA TTTATAAATTTATCTAGGCTGAATTTAACAAGTATTTCCAGAAACCTAAGTTGGCAAAGGCT TTGTTTGAGTATATCTTTCATCATGAAATGATGTTAAGGCTTTGGATCTGGCTGCCCTCTCC ACAGAACATTCTCAGTACAAGGACTGGTGGTGGAAAGTACAGATTGGAATGTTACTACAGG TTGGGAATGTATCGTGAAGCAGAAAAACAGTTTAAATCAGCCCTGAAGCAGCAGGAAATGGTA GATACATTTCTGTACTTGGCAAAAGTATATGTCTCATTGGATCAACCTGTGACTGCTTTAAAT CTTTCAAACAAGGCTTAGATAAGTTTCCAGGAGAAGTAACCCTGCTCTGTGGAATTGCAAGA		

	ATCTATGAGGAAATGAACAATATGTCATCAGCAGCAGAATATTACAAAGAAGTTTGAACAAGACAATACTCATGTGGAAGCCATCGCATGCATTGGAAGCAACCACTTCTATTCTGATCAGCCAGAAATAGCTCTCCGGTTTTACAGGCGGCTGCTGCAGATGGGCATTTATAACGGCCAGCTTTTTAACAACTCTGGGGCTGTGTGCTTCTATGCCCAGCAGTATGATATGACTCTGACCTCATTTGAAACGTGCCCTTTCTTTGGCTGAAAATGAAGAAGAGGCAGCTGATGTCTGGTACAACCTGGGACATGTAGCTGTGGGAATAGGAGATACAAATTTGGCCCATCAGTGCTTCAGGCTGGCTCTGGTCAACAACAACAACCACGCCGAGGCCCTACAACAACCTGGCTGTGCTGGAGATGCCGAAGGGCCACGTTGAACAGGCAAGGGCACTATTACAACTGCATCATCATTAGCACCCCATATGTATGAACCGCATTTAATTTTGCAACAATCTCTGATAAGATTGGAGATCTGCAGAGAAGCTATGTTGCTGCGCAGAAGTCTGAAGCAGCATTTCCAGACCATGTGGACACACAACATTTAATTAAACAATTAAGGCAGCATTTTGCTATGCTCTGATTGTTCTTAGACCACATATGTTCTTATGAAGCAGCATTATGCAAGGGGAAAAAAGCACTATGCTGTGTATGTATGTATATAGTGAATACGTATATTTTAACAACACTGTCCTTGATATTAGTTAAGGTGACACATAAGGGTGAC		
	ORF Start: ATG at 16		ORF Stop: TGA at 1528
	SEQ ID NO: 160	504 aa	MW at 57182.6kD
NOV40a, CG122738-01 Protein Sequence	MSSEMEPLLLAWSYFRRRKFQCADLCTQMLEKSPYDQAAWILKARALTEMVYIDEIDVDQEGIAEMMLDENAIQVPRPGTSLKLPGNQTGGPSQAVRPITQAGRPI TGFLRPSTQSGRPGTMEQAIRTPRTAYTARPITSSSGRFVRLGTASMLTSPDGPFINLSRLNLTKYSQKPKLAKALFEYIFHHENDVKALDLAALSTEHSQYKDWKWKVQIGKCYRRLGMYREAEKQFSAKQOEMVDTFLYLAKVYVSLDQPVLTALNLFKQGLDKFPGEVTL LCGIARIYEEMNNMSSAAEYYKEVLKQDNTHVEAIACIGSNHFYSDQPEIALRFYRRLQMGIFYNGQLFNGLCCFYAQQYDMTLTSFERALSLAENEEEAADVWYNLGHVAVGIGDTNLAHQCFRLALVNNNNHAEAYNNLAVLEMRKGHVQARALLQTASSLAPHMYEPHFNFATISDKIGDLQRSYVAAQKSEAFPDHVDTHQLIKQLRQHFAML		

Further analysis of the NOV40a protein yielded the following properties shown in Table 40B.

Table 40B. Protein Sequence Properties NOV40a	
PSort analysis:	0.5944 probability located in mitochondrial matrix space; 0.3651 probability located in microbody (peroxisome); 0.3022 probability located in mitochondrial inner membrane; 0.3022 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV40a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 40C.

Table 40C. Geneseq Results for NOV40a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG18795	Novel human diagnostic protein #18786 - Homo sapiens, 550 aa. [WO200175067-A2, 11-OCT-2001]	226..504 270..550	254/281 (90%) 262/281 (92%)	e-142
AAM41765	Human polypeptide SEQ ID NO 6696	1..240 20..260	234/241 (97%) 236/241 (97%)	e-133

	[WO200153312-A1, 26-JUL-2001]			
ABG18794	Novel human diagnostic protein #18785 - Homo sapiens, 207 aa. [WO200175067-A2, 11-OCT-2001]	38..243 1..207	191/207 (92%) 197/207 (94%)	e-105
AAB53386	Human colon cancer antigen protein sequence SEQ ID NO:926 - Homo sapiens, 220 aa. [WO200055351-A1, 21-SEP-2000]	339..504 55..220	166/166 (100%) 166/166 (100%)	2e-93
ABG18793	Novel human diagnostic protein #18784 - Homo sapiens, 142 aa. [WO200175067-A2, 11-OCT-2001]	14..154 2..142	141/141 (100%) 141/141 (100%)	1e-76

In a BLAST search of public sequence databases, the NOV40a protein was found to have homology to the proteins shown in the BLASTP data in Table 40D.

Table 40D. Public BLASTP Results for NOV40a

Protein Accession Number	Protein/Organism/Length	NOV40a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8TAM2	Similar to putative - Homo sapiens (Human), 531 aa.	1..504 1..531	504/531 (94%) 504/531 (94%)	0.0
Q9DCP7	0610012F22Rik protein - Mus musculus (Mouse), 505 aa.	1..504 1..505	484/505 (95%) 497/505 (97%)	0.0
Q8VD72	Similar to RIKEN cDNA 0610012F22 gene - Mus musculus (Mouse), 515 aa.	1..504 1..515	484/515 (93%) 497/515 (95%)	0.0
CAD38757	Hypothetical protein - Homo sapiens (Human), 481 aa (fragment).	1..504 7..481	474/505 (93%) 474/505 (93%)	0.0
Q96DG8	Similar to RIKEN cDNA 0610012F22 gene - Homo sapiens (Human), 353 aa (fragment).	153..504 1..353	352/353 (99%) 352/353 (99%)	0.0

5

PFam analysis predicts that the NOV40a protein contains the domains shown in the Table 40E.

Table 40E. Domain Analysis of NOV40a			
Pfam Domain	NOV40a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TPR	214..247	9/34 (26%) 23/34 (68%)	0.92
TPR	281..314	6/34 (18%) 24/34 (71%)	0.019
TPR	349..382	10/34 (29%) 24/34 (71%)	0.0013
TPR	386..419	10/34 (29%) 23/34 (68%)	0.021
TPR	420..453	11/34 (32%) 25/34 (74%)	0.015

Example 41.

The NOV41 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 41A.

5

Table 41A. NOV41 Sequence Analysis			
	SEQ ID NO: 161	3397 bp	
NOV41a, CG123451-01 DNA Sequence	<p>GCGGGACGCGCAGCGCTATGGCAGAGGGCAGCGGGGAAGTGGTCGCAGTGTCTGCGACCGGGG CTGCCAACGGCCTCAACAATGGGGCAGGCGGGACCTCGGCGACGACCTGCAACCCGCTGTCGC GCAAGCTGCATAAGATCCTGGAGACGGGCTGGACAACGACAAGGAGATGTTAGAAGCTCTCA AGGCACTTCAACCTTTTGTGTTGAAAATAGTCTGCGGACTCGAAGAAATTTACGTGGAGATA TTGAACGTAAAAGTTTAGCCATCAATGAAGAATTTGTAAGCATTTTCAAGGAAGTGAAGGAGG AACTTGAAAGCATAAGCGAAGATGTTCAAGCAATGAGCAACTGTTGTCAAGATATGACAAGTC GCCTACAGGCAGCAAAGGAACAGACTCAAGATTTAATAGTTAAAACCACTAAGCTTCAATCTG AAAGCCAAAATTAGAGATAAGAGCTCAAGTTGCAGATGCCTTCTTATCCAAGTTCCAAGTGA CTTCTGATGAAATGAGTCTTCTTCGAGGTACAAGAGAAGGACCCATTACTGAGGATTTTTC AGGCACTGGGAAGAGTAAAACAGATTCATAATGATGTCAAAGTTCTCTTGCGTACAAATCAAC AAACGGCAGGTTTAGAAATTATGGAACAGATGGCCTTACTTCAAGAAACGGCTTATGAAAGAC TTTACCGATGGGCTCAAAGTGAATGCAGAACATTGACACAAGAATCATGTGACGTATCTCCAG TATTGACACAGGCAATGGAAGCCCTGCAGGACAGACCTGTCTTATATAAATATACCTTAGATG AATTTGGAACAGCCAGAAGAAGTACAGTTGTTTCGTGGATTATTGATGCGCTCACAAGAGGGG GCCCCGAGGTACACCTAGACCAATTGAAATGCATCTCATGACCCTTTGAGGTATGTAGGAG ATATGTTGGCTTGGCTCCATCAAGCTACTGCTTCTGAAAAGGAACACCTTGAAGCTCTCTTAA AGCATGTAACACACAAGGTGTTGAAGAAAATATTCAAGAAGTTGTTGGCATATCACTGAAG GTGTGTGCAGGCCCTTAAAGGTTGGAATTGAGCAAGTAATAGTTGCTGAACTGGGCGAGTTT TATTATATAAAATTTCTAATCTCCTCAAATTTATCACCATACAATCAGTGGTATTGTTGGAA ATAGTGCAACTGCATTATGACTACCATTGAAGAAATGCATTTGCTAAGCAAAAAATATTCT TCAATAGCTTGAGTCTTCATGCAAGTAAATTAATGGACAAGGTTGAACTCCCACCACCTGATC TTGGACCAAGTTCTGCATAAATCAGACACTCATGTTGCTGCGTGAAGTTTAGCATCTCAGC ATTCTTCAGTTGTACCATTAGATGCTCGTCAAGCTGATTTTGTGCAAGTTTTATCATGTGTCT TGGATCTCTCTACAGATGTGTACTGTATCAGCCAGCAATTTAGGCACAGCTGACATGGCCA CTTTCATGGTCAATTCATATATATGATGAAGACAACATTAGCTCTATTTGAATTCACTGACA GACGTCTGGAATGCTACAGTTTCAAGTCGAAGCACATTTGGACACACTTATAAATGAGCAAG CCTCTTATGTTTAACTAGGGTAGGCTTGAAGTTACATCTATAACACTGTACAGCAACATAAAC CTGAACAGGGCTCTTAGCTAATATGCCAACCTAGATTCTGTGACACTGAAGGCTGCAATGG TTCAGTTTGATCGTTATCTGTGACCCCGAGACAACCTATTGATACCACAGCTGAACCTTCTTC TAAGTGCCACAGTGAAGAGCAGATCGTAAAACAATCTACAGAATTAGTCTGCAGAGCCTATG</p>		

	GTGAAGTGTATGCAGCCGTGATGAATCCAATCAATGAATACAAAGATCCAGAGAACATTCTTC ACCGATCGCCGCAGCAAGTGCAGACGCTTCTTTCTGATTATCTTATTTTCATTGTGTTAGCAA AATATGACCTCCCTAAAACACTGAAGGTTATTTTTTATTCTTTGAATTTTACTTTATAATTT GATAGTTACAGTTTTCTTTGTATCATAAGATTGTAAGTCCCGATAATTTTTTTTTTTTGGTC TCAGTAACAGGGAAGTAAGTAACATGTTGACCTGAGCTAGTATTGCTGTGTATCTACTCTAAA TGAGATGATCTATTTTTTTGCTAGCCATCTCTCCAGCTCTGCAGTTTTCACTGTATTCAGGAA GCATAAAGTAGTATGAAAGGTTTGAAGAATTTTTTTTACAAGACTAGTTCTAAATTAACAGC TTATAAAAAATTTGTCTAAATTTAATAATTAGTATAAGGATATGACCTAATAAATGTCTCCTT ACCTAAAGATTCATTGCTTTCTTTTAAATATGAGTAGGCATACTTAGTAGCTTTTCTGAACCT AGCCTATGCTCTGTGTCCTCCAAAATAGCTGCCCTTAAAGAGTTGTTAGCAGAGAGAAAAATAAC AGTGAATGTGCTCCTGGTGTATATGGCAGTGAATCTCCTTTTCTGTTCTACTTTAGCATACTAT ATATATTTGACTGTGTACATTCTTATGCAATTTTAAAGTATACACTCAGCAATAATTAGAAAAA AAGGAGAGAGAAAAAGTGATTTAAACAGGGTGGATTCCACTCTGTGGGAGCCTTCGATGGAAC CAAGGTGGAGCTCAGCCTTTCCAATGAGCTCTAAGCATGTAGATAGCCTGAGCTGTGTCTAAG CCTGGTGTTTAAAGATGGGTATTTGTATACAATATGGGTCTTAAATCCAACCAACTACACAT TTTATCTGGTGTTCAAACCAAAGAAACAATGATCTACTCAAACATTGGAGAAAAAACTGCCA GAGGAGGAGTTGCCAATTGGCAGTGTGTCTTATCTCCATGTTGTAACCTGGACTCTGACTTTAG ACCATTACCTATTAGGAAGATTAAAAATGACTGTATTTTTAAAGGAATAAATCCCAGTGTGCC TGATTTGACATTCTTGTGCAGAAAAAACTTAATTTCTAGTAAATCTATAAAATGGGTAAG TCCCTAAATTACAAATGAGAAAATTGAAGCACAAGGAAAAAAATAACTAGTTTGAAATATTTT GAAAAGTAATAACATAAACTAGTATTTGTAGAAGATTATGTGTTGTATATAACAAATTAGTA TTTATAGAATATGACCTATTTATCTGAAGTTTATAATTGTTTATACCTAATACAGTTCTTTTT GGAGTAAGAATGATTATATAATCGTTATCCATTGGGTATAAATCTGTATTTTGTGTTTTTC CCTTTGATTAGTATGTGTTACATATAAAGACAGAAAAATAAGTATAAATCAAGAGCTT		
	ORF Start: ATG at 18		ORF Stop: TGA at 1989
	SEQ ID NO: 162	657 aa	MW at 73278.2kD
NOV41a, CG123451-01 Protein Sequence	MAEGSGEVVAVSATGAANGLNNGAGGTSATTCNPLSRKLHKILETRLDNDKEMLEALKALSTF FVENSLRTRNRNRGDIERKSLAINEEFVSIKFEVKEELESISEDVQAMSNCODMTSRLQAAK EQTQDLIVKTTKLQSESQKLEIRAQVADAFLSKFQLTSDMSLLRGTRREGPITEDFFKALGRV KQIHNDVKVLLRTNQQTAGLEIMEQMALLQETAYERLYRWAQSECRILTQESCDVSPVLTQAM EALQDRPVLYKYTLDEFGTARRSTVVRGFDALTRGGPGGTPRPIEMHSHDPLRYVGDMLAWL HQATASEKEHLEALLKHVTTQGVENIQEVVGHITGVCRPLKVRIEQVIVAEPGAVLLYKIS NLLKFYHHTISGIVGNSATALLTTEEMHLLSKKIFFNSLSLHASKLMDKVELPPPDLGPSSA LNQTLMLLREVLASHDSSVVPLDARQADFVQVLSCLVDPLLMQCTVSASNLGTADMATFMVNS LYMMKTTLLALFEFTDRRLEMLQFQIEAHLDTLINEQASYVLTRVGLSYIYNTVQQHKEPQGS ANMPNLDSVTLKAAMVQFDRLSAPDNLIPQLNPLLSATVKEQIVKQSTELVCRAVGEVYAA VMNPINEYKDPENILHRSPPQVQTLLS		

Further analysis of the NOV41a protein yielded the following properties shown in Table 41B.

Table 41B. Protein Sequence Properties NOV41a	
PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in lysosome (lumen); 0.1000 probability located in endoplasmic reticulum (lumen); 0.1000 probability located in outside
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV41a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 41C.

Table 41C. Geneseq Results for NOV41a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM41398	Human polypeptide SEQ ID NO 6329 - Homo sapiens, 670 aa. [WO200153312-A1, 26-JUL-2001]	1..657 14..670	655/657 (99%) 655/657 (99%)	0.0
AAM39612	Human polypeptide SEQ ID NO 2757 - Homo sapiens, 656 aa. [WO200153312-A1, 26-JUL-2001]	1..657 1..656	653/657 (99%) 653/657 (99%)	0.0
ABB58712	Drosophila melanogaster polypeptide SEQ ID NO 2928 - Drosophila melanogaster, 630 aa. [WO200171042-A2, 27-SEP-2001]	35..657 14..628	262/632 (41%) 411/632 (64%)	e-147
ABP04367	Human ORFX protein sequence SEQ ID NO:8716 - Homo sapiens, 122 aa. [WO200192523-A2, 06-DEC-2001]	516..636 1..121	109/121 (90%) 118/121 (97%)	2e-57
AAB41840	Human ORFX ORF1604 polypeptide sequence SEQ ID NO:3208 - Homo sapiens, 107 aa. [WO200058473-A2, 05-OCT-2000]	53..159 1..107	106/107 (99%) 106/107 (99%)	4e-51

In a BLAST search of public sequence databases, the NOV41a protein was found to have homology to the proteins shown in the BLASTP data in Table 41D.

Table 41D. Public BLASTP Results for NOV41a

Protein Accession Number	Protein/Organism/Length	NOV41a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y2V7	Hypothetical 68.1 kDa protein - Homo sapiens (Human), 605 aa.	53..657 1..605	605/605 (100%) 605/605 (100%)	0.0
Q9ULT5	KIAA1134 protein - Homo sapiens (Human), 611 aa (fragment).	5..609 1..605	605/605 (100%) 605/605 (100%)	0.0
Q8R3I3	Similar to KIAA1134 protein - Mus musculus (Mouse), 605 aa.	53..657 1..605	559/605 (92%) 587/605 (96%)	0.0
Q9V564	CG1968 protein - Drosophila melanogaster (Fruit fly), 630 aa.	35..657 14..628	262/632 (41%) 411/632 (64%)	e-147
Q9C6R8	Hypothetical 78.8 kDa protein - Arabidopsis thaliana (Mouse-ear cress), 706 aa.	35..656 11..704	236/707 (33%) 375/707 (52%)	e-103

Pfam analysis predicts that the NOV41a protein contains the domains shown in the Table 41E.

Table 41E. Domain Analysis of NOV41a			
Pfam Domain	NOV41a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5 Example 42.

The NOV42 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 42A.

Table 42A. NOV42 Sequence Analysis			
	SEQ ID NO: 163	3779 bp	
NOV42a, CG123660-01 DNA Sequence	GCGGCCGCCCCGGCTCCCGCTGCAGGAATCGCGCCAGGACGCTGGCCCCGCTCGCGGCTAGC TTGCACGCCAGGGCACAGCGAGGATGGGAGGGTCGCAGTCCCTGCAGCCAGCCCCAGCCAGCG ACCTGAACCTGGAGGCTTCCGAGGCAATGAGTTCCGATTCTGAAGAGGCATTGAGACCCCGG AGTCAACGACCCCTGTCAAAGCTCCGCCAGCTCCACCCCCACCACCCCGAAGTCATCCCAG AACCCGAGGTCAGCACACAGCCACCCCGGAAGAACCAGGATGTGGTCTCGAGACAGTCCCTG TCCCTGATGGGCCACGGAGCGACTCGGTGGAAGGAAGTCCCTTCCGTCCCCCGTCACACTCCT TCTCTGCCGTCTTCGATGAAGACAAGCCGATAGCCAGCAGTGGGACTTACAACCTGGACTTTG ACAACATTTAGGCTTGTGGATACCTTTCAGACCTTGAGGCTCGTGCCTCAGACGCTAAGAATC AGGAGGGCAAAGTGAACACACGGAGGAAGTCCACGGATTCCGTCCCCATCTTAAGTCTACAC TGTCCCGGTTCGCTCAGCCTGCAAGCCAGTGACTTTGATGGTGCTTCTCCTCAGGCAATCCCG AGGCCGTGGCCCTTGCCCCAGATGCATATAGCACGGGTTCCAGCAGTGCTTCTAGTACCCTTA AGCGAACTAAAAACCGAGGCCGCTTCTTAAAAAGAAACAGACCACCAAGAAACCCACAG AGACCCCCCAGTGAAGGAGACGCAACAGGAGCCAGATGAAGAGAGCCTTGTCCCAGTGGGG AGAATCTAGCATCTGAGACGAAACCGGAATCTGCCAAGACGGAAGTCCAGCCAGCCTTAT TGGAGGAGACGCCCTTGAGCCCGCTGTGGGGCCCAAAGCTGCCTGCCCTCTGGACTCAGAGA GTGCAGAAGGGTTGTCCCCCGCTTCTGGAGGTGGCAGAGTGCAGAACTACCCCTGTCTG GGAGGAAAACGCTGCCTCTTACCACGGCCCGGAGGCAGGGGAGGTAACCCCATCGGATAGCG GGGGCAAGAGGACTCTCCAGCCAAAGGCTCTCCGTAAGGCTGGAGTTTGACTATTCTGAGG ACAAGAGTAGTTGGGACAACAGCAGGAAAACCCCTCTTACCAAAAAGATAGGCAAAAAGC CAGTTGCCAAATGCCCTGAGGAGGCCAAAGATGAAAAGACACCCGAGAACTTGACAACA CTCCTGCCTCAGCTCCAGATCCCTGTGTAACCAATGACATCCCATTGTAAAGGTACTT ACACCTTTGATATTGACAAGTGGGATGACCCCAATTTTAAACCTTTTCTTCCACCTCAAAAA TGCAGGAGTCTCCAAACTGCCCAACATCATACAACTTGACCCAGACACCTGTGATGAGT CCGTTGACCCCTTTAAGACATCCTCTAAGACCCCGAGCTCACCTTCTAAATCCCCAGCCTCCT TTGAGATCCCAGCCAGTGCTATGAAGCCAATGGAGTGGACGGGATGGGCTAAACAAGCCCG CCAAGAAGAAGACGCCCTTAAAGACTGACACATTTAGGGTGAAAAAGTCGCCAAAACGGT CTCCTCTCTGATCCACCTTCCAGGACCCACCCAGCTGCTACACGAGAAACACCAACGAG TGATCTCTGCGTGGTCCAGCCACAGATGAGGAAAAGCTGGCGGTCAACCAACGAGAAGTGA TGTGCATGACAGTGGACCTAGAGGCTGACAAACAGGACTACCCGAGCCCTCGGACCTGTCCA CCTTTGTAAACGAGACCAATTCAGTTACCCACTGAGGGTAAGCAACTGTGTAGCCAGCTGG ACCCCACTCTGCCTCGGAGAATCCTGCCCTAGGGAGCCAAAGGCCAGGAGAGAACTTCTC AGCCAGAGTTGGATTACAGAACTCCTATGAAATGAATATATGGAGAAAATGGCTCCTCCT TACCTCAGGACGACGATGCCCCGAAGAAGCAGGCCTTGTACCTTATGTTGACACTTCTCAGG AGAGCCCTGTCAAGTCATCTCCGTCCGCATGTCAGAGTCCCCGACGCCGTGTTACGGGTCAA GTTTGAAGAGACTGAAGCCCTTGTGAACACTGCTGCGAAAAACAGCATCCTGTCCCACGAG GACTGGCCCTTAACCAAGAGTCACTTGCAGGTGCCAGAGAAATCCTCCAGAGGAGCTGG AGGCCATGGGCTTGGGCACCCCTTCAGAAGCGATTGAAATTACAGTCCCGAGGGCTCCTTTG CCTCTGCTGACGCCCTCTCAGCAGGCTAGCTCACCCTGCTCTCTGTGGTGCCTTGACT ATCTGGAGCCCGACTTAGCAGAAAAGAACCCCACTATTCGCTCAGAACTCCAGGAGGAGT		

	<p>TAGAGTTTGCCATCATGCGGATAGAAGCCCTGAAGCTGGCCAGGCAGATTGCTTTGGCTTCCC GCAGCCACCAGGATGCCAAGAGAGAGGCTGCTACCCCAACAGACGTCTCCATCTCCAAAACAG CCTTGTA TCTCCCGCATCGGGACCGCTGAGGTGGAGAACTGCAGGCCTTCTGTTCCAGCAGC CCGACCTGGACTCTGCCCTCCAGATCGCCAGAGCAGAGATCATAACCAAGGAGAGAGAGGTCT CAGAATGGAAGATAAATATGAAGAAAGCAGGCGGGAAGTGATGGAAATGAGGAAAATAGTGG CCGAGTATGAGAAGACCATCGCTCAGATGATAGAGGACGAACAGAGAGAGAAGTCACTCTCCC ACCAGACGGTGACAGCAGCTGGTTCGAGAGAAGGAGCAAGCCCTGGCCGACCTGAAGTCCGTGG AGAAGTCTCTGGCCGACCTCTCAGAAGATATGAGAAGATGAAGGAGTCTTAGAAGGCTTCC GCAAGAATGAAGAGGTGTTGAAGAGATGTGCGCAGGAGTACCTGTCCCGGTGAAGAAGGAGG AGCAGAGGTACCAGGCCCTGAAGGTGCACGCGGAGGAGAACTGGACAGGGCCAATGCTGAGA TTGCTCAGGTTGAGGCAAGGCCAGCAGGAGCAAGCCGCCACCAGGCCAGCTGCGGAAGG AGCAGCTGCGAGTGGACGCCCTGGAAGGACGCTGGAGCAGAAGAATAAAGAAATAGAAGAAC TCACCAAGATTTGTGACGAAGTGTGCCCCAAATGGGGAAGCTAACTCTGAACCGAATGTT TTGGACTTAAGTGTGCGGCAATATGACCGTCGGCACACTGCTGTTCTCCAGTTCATGGAC AGGTTCTGTTTTCACCTTTTTCGTATGCACTACTGTATTTCTTCTAAATAAAATGATTGTA TTGTATGCACTACTAAGGAGACTATCAGAAATTTCTTGCTATTGGTTTGCATTTCTCTAGTATA ATTCATAGCAAGTTGACCTCAGAGTCTCTGTATCAGGGAGATTGTCTGATTCTCTAATAAAAG ACACATTGCTGACCTTGGCCTTGGCCTTTGTACACAAGTTCCAGGGTGAGCAGCTTTTGGAT TTAATATGAACATGTACAGCGTGCATAGGGACTCTTGCTTAAAGGAGTGTAAGTCTGATCTGC ATTGCTGATTTGTTTTTAAAAAACAAGAAATGCATGTTTCAAATAAAATTTCTATTGTAA ATAAAATTTTCTTGGATCTTGAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG</p>
	<p>ORF Start: ATG at 87</p> <p>ORF Stop: TAA at 3258</p>
	<p>SEQ ID NO: 164</p> <p>1057 aa</p> <p>MW at 115537.2kD</p>
NOV42a, CG123660-01 Protein Sequence	<p>MGGSQLQPAPASDLNLEASAMSSDSEEFETPESTTPVKAPPAPPPPPPEVIPEPEVSTQP PPEPGCGSETVPVDPGRSDSVEGSPFRPPSHSFAVFDEDKPIASSGTYNLDFDNIELVDT FQLEPRASDAKNQEGKVNTRKSTDSPISKSTLSRSLQASDFDGAASSGNPEAVALAPD AYSTGSSASSTLKRKKPRPPSLKKQTTKKPTETPPVKETQEPDEESLVPGENLASETK TESAKTEGSPALLEETPLEPAVGPKAACPLDSESAEGVVPASGGGRVQNSPPVGRKTLPLT TAPEAGEVTPSDSGGQEDSPA KGLSVRLEFDYSEDKSSWDNQENPPPTKKIGKKPVAKMPLR RPMKKTPEKLDNTPASPPRSPAEPNDIPIAKGTYTFDIDKWDDPNFNPFSSTSKMQESP KPL QQSYNFDPTCDESVDPFKTSSTPSSPSKSPASFEIPASAMEANGVDGDGLNPKAKKKKTLPL KTDTFRVKSPKRSPLSDPPSQDPTPAATPETPPVISAVVHATDEKLAVTNQKWMCMCTVDLE ADKQDYPQPSDLSTFVNETKFSSPTGKQLCSQLDPHSASENPAPREP KARRETSQPELDYRN SYEIEYMEKIGSSLPQDDAPKQALYLMFDTQSQSPVKSSPVRMSSEPTPCSGSSFEETEAL VNTAAKNQHPVPRGLAPNQESHQVPEKSSQKELEAMGLGTPSEAIETAPEGSFASADALLS RLAHPVSLCGALDYLEPDLAEKNPPLFAQKLQEELEFAIMRIEALKLARQIALASRHQDAKR EAAHPTDVSISKALYSRIGTAEVEKPAGLLFQQPDLSALQIARAEIITKEREVSEWKDKYE ESRREVMEMRKIVAEYEKTIQMIEDEQREKSVSHQTVQQLVLEKEQALADLNSVEKSLADLF RRYEKMEVLEGRFKNKEVLKRCQAEYLSRVKKEEQRYQALKVHAEKLDNRANAIEAQVRGKA QQEQAAHQASLRKEQLRVDALERTLEQKNKEIEELTKICDELIAMGKS</p>
	<p>SEQ ID NO: 165</p> <p>3686 bp</p>
NOV42b, CG123660-02 DNA Sequence	<p>GCGGCCGCCCCGCTCCCGCTGCAGGAATCGCGCCAGGACGCTGGCCCCGCTCGCGGCTAGC TTGCACGCCAGGGCACAGCGAGGATGGGAGGGTCGCAGTCCCTGCAGCCAGCCCCAGCCAGCG ACCTGAACCTGGAGGCTTCCGAGGCAATGAGTTCGATTCTGAAGAGGCATTGAGACCCCCG AGTCAACGACCCCTGTCAAAGCTCCGCCAGCTCCACCCCCACCACCCCCGAAGTCATCCCAG AACCCGAGGTGAGCACACAGCCACCCCGGAAGAACCAGGATGTGGTTCTGAGACAGTCCCTG TCCCTGATGGCCACGAGCGACTCGGTGGAAGGAAGTCCCTCCGTCCCCCGTCACACTCCT TCTCTGCCGTCTTCGATGAAGACAAGCCGATAGCCAGCAGTGGGACTTACAACCTGGACTTTG ACAACATTGAGCTTGTGATACCTTTTACAGCCTTGGAGCCTCGTGCCTCAGACGCTAAGAATC AGGAGGGCAAAGTGAACACACGAGGAGGAGTCCACGGATTCCGTCCCCATCTTAAGTCTACAC TGTCCCGGTCGCTCAGCCTGCAAGCCAGTGACTTTGATGGTGCTTCTCTCTCAGGCAATCCCG AGGCCGTGGCCTTGGCCAGATCATATAGCACGGGTCCAGCAGTGCTTCTAGTACCTTA AGCGAACTAAAAACCGAGGCGCCTTCTTAAAAAAGAAACAGACCACCAAGAAACCCACAG AGACCCCCCAGTGAAGGAGACGCAACAGGAGCCAGATGAAGAGAGCCTTGTCCCAGTGGGG AGAATCTAGCATCTGAGACGAAAACGGAATCTGCCAAGACGGAAGGTCTAGCCCAGCCTTAT TGGAGGAGACGCCCCTTGAGCCCGCTGTGGGGCCCAAAGCTGCCTGCCCTCTGGACTCAGAGA GTGCAGAAGGGGTTGTCCCCCGGCTTCTGGAGGTGGCAGAGTGCAGAAGTCAACCCCTGTG GGAGGAAAACGCTGCCTTTACACGCGCCCCGAGGAGGAGGTAACCCCATCGGATAGCG GGGGGCAAGAGGACTCTCCAGCCAAAGGGCTCTCCGTAAAGGCTGGAGTTTGAATTTCTGAGG ACAAGAGTAGTTGGGACAACAGCAGGAAAACCCCTCTACCAAAAAGATAGGCAAAAAGC CAGTTGCCAAAATGCCCTGAGGAGGCCAAAGATGAAAAGACCCGAGAACTTGACAACA CTCCTGCCTCACCTCCAGATCCCCTGCTGAACCAATGACATCCCCATTGCTAAAGGTACTT ACACCTTTGATATTGACAAGTGGGATGACCCCAATTTTAAACCTTTTCTTCCACCTCAAAA TGCAGGAGTCTCCAAACTGCCCCAACCAATCATACAACCTTTGACCCAGACACCTGTGATGAGT</p>

	CCGTTGACCCCTTTAAGACATCCTCTAAGACCCCCAGCTCACCTTCTAAATCCCCAGCCTCCT TTGAGATCCCAGCCAGTGCTATGGAAGCCAATGGAGTGGACGGGGATGGGCTAAACAAGCCCG CCAAGAAGAAGACGCCCTTAAAGACTGACACATTTAGGGTGAAAAGTCGCCAAAACGGT CTCCTCTCTGATCCACCTTCCCAGGACCCACCCAGCTGCTACACCAGAAAACACCACAG TGATCTCTGCGGTGGTCCACGCCACAGATGAGGAAAAGCTGGCGGTACCAACCAGAAGTGGA CGTGATGACAGTGGACCTAGAGGCTGACAAACAGGACTACCCGCAGCCCTCGGACCTGTCCA CCTTTGTAACGAGACCAAATTCAGTTCACCCACTGAGGAGTTGGATTACAGAAACTCCTATG AAATTGAATATATGGAGAAAATTGGCTCCTCCTTACCTCAGGACGACGATGCCCCGAAGAAGC AGGCCTTGACCTTATGTTTGACACTTCTCAGGAGAGCCCTGTCAAGTCATCTCCCGTCCGCA TGTCAGAGTCCCCGACGCCGTGTTTCAAGGTCAAGTTTGAAGAGACTGAAGCCCTTGTGAACA CTGCTGCGAAAAACAGCATCCTGTCCCACGAGGACTGGCCCTTAACCAAGAGTCACACTTGC AGGTGCCAGAGAAATCCTCCAGAAGGAGCTGGAGCCATGGGCTTGGGCACCCCTTCAGAAG CGATTGAAATTACAGCTCCGAGGGCTCCTTGCCTCTGCTGACGCCCTCCTCAGCAGGCTAG CTCACCCCGTCTCTCTGTGGTGAAGTACTATCTGGAGCCGACTTAGCAGAAAAGAACCC CCCCCTATTGCTCAGAACTCCAGGAGGAGTTAGAGTTTGCCATCATGCGGATAGAAGCCC TGAAGCTGGCCAGGCAGATTGCTTTGGCTTCCCGCAGCCACCAGGATGCCAAGAGAGAGGCTG CTCACCAACAGACGTCTCCATCTCCAAAACAGCCTTGTAATCCCGCATCGGACCGCTGAGG TGGAGAAACCTGCAGGCCTTCTGTTCCAGCAGCCCGACCTGGACTTGCCTCCAGATCGCCA GAGCAGAGATCATAACCAAGGAGAGAGAGGTCTCAGAATGGAAAGATAAATATGAAGAAAGCA GGCGGGAAGTGATGGAATGAGGAAAATAGTGGCCGAGTATGAGAAGACCATCGCTCAGATGA TAGAGGACGAACAGAGAGAGAAGTCAGTCTCCACAGACGCTGCAGCAGCTGGTTCTGGAGA AGGAGCAAGCCCTGGCCGACCTGAACTCCGTGGAGAAGTCTTGGCCGACCTCTCAGAAGAT ATGAGAAGATGAAGGAGGTCTAGAAGGCTTCCGCAAGAATGAAGAGGTGTTGAAGAGATGTG CGCAGGAGTACCTGTCCCGGTGAAGAAGGAGGAGCAGAGTACCAGGCCCTGAAGGTGCACG CGGAGGAGAACTGGACAGGGCCAATGCTGAGATTGCTCAGGTTTCGAGGCAAGGCCACGAGG AGCAAGCCGCCACCAGGCCAGCCTGCGGAAGGAGCAGCTGCGAGTGGACGCCCTGGAAGGA CGCTGGAGCAGAAGAATAAAGAAATAGAAGAACTACCAAGATTTGTGACGAATGATTGCCA AAATGGGGAAGCTAACTCTGAACCGAATGTTTGGACTTAAGTGTGGCGCAATATGACCG TCGGCACACTGCTGTTCCCTCAGTTCCATGGACAGGTTCTGTTTCACTTTTCGTATGCACT ACTGTATTTCTTTCTAAATAAAATTGATTGATTGTATGCAGTACTAAGGAGACTATCAGAA TTTCTTGCTATTGGTTTGGATTTTCTAGTATAATTCATAGCAAGTTGACCTCAGAGTCTCTG TATCAGGAGATTGTCTGATTCTCTAATAAAGACACATTGCTGACCTTGGCCTTGGCCTTTG TACACAAGTTCCAGGGTGAAGCAGCTTTTGGATTAAATATGAACATGTACAGCGTCATAGGG ACTCTTGCTTAAAGGAGTGAACCTTGATCTGCATTGCTGATTTGTTTTAAAAAACAAGA AATGCATGTTTCAAATAAAATTCTCTATTGTAATAAAATTTTTCTTTGGATCTTGAAAAAA AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 87 ORF Stop: TAA at 3165		
	SEQ ID NO: 166	1026 aa	MW at 112109.5kD
NOV42b, CG123660-02 Protein Sequence	MGGSQSLQFAPASDLNLEASEAMSSDSEAFETPESTTPVKAPPAPPPPPPEVIPEPEVSTQP PPEEPGCGSETVVPDGPDSVEGSPFRPPSHSFAVFDKPIASSGTYNLDFDNI ELVDT FQTLERPRASDAKNEQGVNTRRKSTDSVPISKSTLSRSLSLQASDFDASSGNPEAVALAPD AYSTGSSASSTLKRTRKPRPPSLKKKQTKKPTETPPVKETQEPDEESLVPSENLASETK TESAKTEGSPALLEETPLEPAVGPKAACPLDSESAEGVPPASGGGRVQNSPPVGRKTLPLT TAPEAGEVTPSDSGGQEDSPAKGLSVRLEFDYSEDKSSWDNQENPPPTKKIGKKPVAKMPLR RPKMKKTPEKLDNTPASPPRSPAEPNDIPIAKGTYTFDIDKWDNPNFNPFSSTSKMQESPKLP QQSYNFDPDTCDESVDPFKTSSTPSSPSKSPASFEIPASAMEANGVDGDLNPKAKKKKTPL KTDTFRVKSPKRSPLSDPPSQDPTPAATPETPPVISAVVHATDEEKLAVTNQKWTCTVDLE ADKQDYPQPSDLSTFVNETHFSSPTEELDYRNSYEIEYMEKIGSSLPQDDAPKKQALYLMFD TSQESPVKSSPVRMSESPTPCGSSFEETEALVNATAKNQHPVPRGLAPNQESHQVPEKSSQ KELEAMGLGTPSEAEITAEPEGSFASADALLSRLAHPVSLCGALDYLEDLAENPNPLFAQKL QEELEFAIMRIEALKLARQIALASRSHQDAKREAAHPTDVSISKALYSRIGTAEVEKPAGLL FQPPDLDSALQIARAEIITKEREVSEWKDYEESRREVMEMRKIVAEYEKTIQMIEDEQREK SVSHQTVQQLVLEKEQALADLNSVEKSLADLFRRYEKMKEVLEGRFKNEEVLKRCQAQEYLSRV KKEEQRYQALKVHAEEKLDRANAIEIAQVRGKAQQEQAAHQASLRKEQLRVDALERTLEQKNKE IEELTKICDELIAMGKS		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 42B.

Table 42B. Comparison of NOV42a against NOV42b.

Protein Sequence	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV42b	1..1057 1..1026	916/1057 (86%) 916/1057 (86%)

Further analysis of the NOV42a protein yielded the following properties shown in Table 42C.

Table 42C. Protein Sequence Properties NOV42a

PSort analysis:	0.9800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV42a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 42D.

Table 42D. Geneseq Results for NOV42a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM38678	Human polypeptide SEQ ID NO 1823 - Homo sapiens, 1013 aa. [WO200153312-A1, 26-JUL-2001]	24..1057 41..1013	972/1034 (94%) 972/1034 (94%)	0.0
AAM38680	Human polypeptide SEQ ID NO 1825 - Homo sapiens, 1025 aa. [WO200153312-A1, 26-JUL-2001]	24..1057 41..1025	972/1046 (92%) 972/1046 (92%)	0.0
AAM40466	Human polypeptide SEQ ID NO 5397 - Homo sapiens, 865 aa. [WO200153312-A1, 26-JUL-2001]	133..1057 1..865	853/926 (92%) 855/926 (92%)	0.0
AAM40465	Human polypeptide SEQ ID NO 5396 - Homo sapiens, 865 aa. [WO200153312-A1, 26-JUL-2001]	133..1057 1..865	853/926 (92%) 855/926 (92%)	0.0
AAM40464	Human polypeptide SEQ ID NO 5395 - Homo sapiens, 865 aa. [WO200153312-A1, 26-JUL-2001]	133..1057 1..865	853/926 (92%) 855/926 (92%)	0.0

10

In a BLAST search of public sequence databases, the NOV42a protein was found to have homology to the proteins shown in the BLASTP data in Table 42E.

Table 42E. Public BLASTP Results for NOV42a

Protein Accession Number	Protein/Organism/Length	NOV42a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O95359	Transforming acidic coiled-coil-containing protein 2 (Anti Zuai-1) (AZU-1) - Homo sapiens (Human), 1026 aa.	1..1057 1..1026	1025/1057 (96%) 1025/1057 (96%)	0.0
Q9JJG0	Transforming acidic coiled-coil-containing protein 2 - Mus musculus (Mouse), 1035 aa.	1..1057 1..1035	886/1071 (82%) 928/1071 (85%)	0.0
Q8TCK9	Hypothetical 88.9 kDa protein - Homo sapiens (Human), 807 aa (fragment).	143..1057 1..807	715/918 (77%) 742/918 (79%)	0.0
Q8WVR1	Hypothetical 64.7 kDa protein - Homo sapiens (Human), 575 aa.	375..1057 1..575	484/687 (70%) 510/687 (73%)	0.0
Q99KQ6	Similar to transforming, acidic coiled-coil containing protein 2 - Mus musculus (Mouse), 598 aa.	375..1057 1..598	452/686 (65%) 497/686 (71%)	0.0

- 5 Pfam analysis predicts that the NOV42a protein contains the domains shown in the Table 42F.

Table 42F. Domain Analysis of NOV42a

Pfam Domain	NOV42a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 43.

- 10 The NOV43 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 43A.

Table 43A. NOV43 Sequence Analysis

	SEQ ID NO: 167	3351 bp	
NOV43a,	GACTCCTGCCTCAGGATGCCGGGGAGGAAGAGGAGCGGGCCTTCCTGGTGGCCCGCGAGGAG		

CG123955-01 DNA Sequence	CTGGCGAGCGCCCTGAGGAGGGATTCCGGGCAGGCGTTTTCCTGGAGCAGCTCCGGCCGCTA CTAGCCAGCTCTCTGCCGCTAGCCGCCGCTACCTGCAGCTGGACGCCGACGCCTTGTCCGC TGCAACGCTCATGGGGAGCCCCGAAACTACCTCAACACCCTGTCCACGGCTCTGAACATCCTG GAGAAATACGGCCGCAACCTTCTCAGCCCTCAGCGGCTCGGTACTGGCGTGGTGTCAAGTTT AATAACCCTGTCTTTTCGACGACGGTGGATGCTGTGACAGCAGGGGGGCGAGATGTGCTGCGA TTATATGGCTACACAGAGGAGCAACCAGATGGGTTGAGCTTCCCCGAAGGGCAGGAGGACCA GATGAGCACCAGGTTGCTACAGTCACTGGAAGTACTGCTGCTTCGGACAGAGCTCAGCCTG CTATTGCAGAATACTCATCCAAGACAGCAGGCACTGGAGCAGCTGTTGGAAGACAAGGTTGAA GATATGCTGCAGCTTTCAGAATTGACCCCCCTATTGAGAGAGATTGCTCCTGGCCCCCTCACC ACACCCTCTGTCCCAGGCTCCACTCCTGGTCCCTGCTTCTCTGTGGTTCTGCCCCAGGCACA CTGCACTGCCCATCCTGTAACAGGCCCTGTGTCCAGCCTGTGACCACCTGTTCCATGGACA CCATCCCGTGCTCATCACCTCCGCCAGACCCTGCCTGGGGTCTGCAAGGTACCCACCTGAG AGTTTACCTGCCTCAGCCCAACCACGGCCCCAGTCGACCTCCCTGCTGGCCCTGGGAGACAGC TCTCTTTCTTCCCCTAATCCTGCAAGTGCTCATTGGCCCTGGCACTGTGCTGCTGTGCCATG CTAAATGAGCCTTGGGCAGTGCTCTGTGTGGCCTGTGATCGGCCCGAGGCTGTAAGGGGTTG GGGTTGGGAAGTGGGGTCCCCAAGGAACTGGAGGCCCTAGAACCTGATCTTGACAGGGGTCCG TGGGCTGCCAGAGCTGTACCTTTGAGAATGAGGCAGCTGCTGTGCTATGTTCCATATGTGAG CGACCTCGGCTGGCCAGCCTCCAGCTTGGTGGTGGATTCCGAGATGCTGGCATTGCTCCTG CAACCCTTTCAGCAGGGGATGCTTTGCTGGCCTCTGCCAGAGTCAAGTCTGCTACTGTATT CACTGTACCTTCTGCAACTCGAGCCCTGGCTGGGTGTGTGTTATGTGCAACCGGACTAGTAGC CCCATTCCAGCACAACATGCCCCCGGCCCTATGCCAGCTCTTTGAAAAGGGACCCCCAAG CCTGGGCCCCACGACGCCTTAGTGCCCCCTGCCAGTTCCTGTGGAGATCCTGAGAAGCAG CGCCAAGACAAGATGCGGAAGAAGGCCCTCCAGCTAGTGAGCATGATCCGGGAAGGGGAAGCC GCAGGTGCCTGTCCAGAGGAGATCTTCTCGGCTCTGCAGTACTCGGGCACTGAGGTGCCTCTG CAGTGGTTGCGCTCAGAACTGCCCTACGTCCTGGAGATGGTGGCTGAGCTGGCTGGACAGCAG GACCTTGGGCTGGGTGCCTTTTCTGTCTCAGGAGGCCCGGAGAGCCTGGCTGGATCGTCATGGC AACCTTGATGAAGCTGTGGAGGAGTGTGTGAGGACCAGGCGAAGGAAGGTACAGGAGCTCCAG TCTCTAGGCTTTGGGCTGAGGAGGGGTCTCTCCAGCGATTGTTCCAGCAGAGGATGATGTG TCACGGGCCCTGACTGAGCTACAGCGCCAACGCCCTAGAGCCCTTCGCCAGCGCCTCTGGGAC AGTGGCCCTGAGCCACCCCTTCTGGGATGGGCCAGACAAGCAGAGCCTGGTCAGCGGCTT TTGGCAGTCTACGCACTCCCAGCTGGGGCCGGGCAGAGCTGGCACTGTCACTGCTGCAGGAG ACACCCAGGAATCATGAGTTGGGGGATGTGGTAGAAGCTGTGAGGCACAGCCAGGACCGGGCC TTCTTGCGCCGCTTGCTTGCCAGGAGTGTGCCGTGTGTGGCTGGGCCCTGCCCCACAACCGG ATGCAGGCCCTGACTTCTGTGAGTGACCATCTGTCTGACTGCTTCCGCCAGCACTTCACC ATCGCCTTGAAGGAGAAGCACATCACAGACATGGTGTGCCCTGCCTGTGGCCGCCCGACCTC ACCGATGACACACAGTTGCTCAGCTACTTCTCTACCTTGACATCCAGTCTCGCGAGAGCCTA GAGCCAGATGCCTATGCGTTGTTCCATAAGAAGCTGACGAGGGTGTGCTGATGCGGGACCCC AAGTTCTTGTGGTGTGCCAGTGCTCCTTTGGCTTCATATATGAGCGTGAGCAGCTGGAGGCA ACTTGTCCCAGTGTCAACGACCTTCTGTGTGCGCTGCAAGCGCCAGTGTGAGGACTTCCAG AACTGGAACCGCATGAACGACCCAGAATACCAGGCCCAGGGCCTAGCAATGTATCTTCAGGAA AACCGCATTGACTGTGCCCAATGCAAGTTCTGTACGCCCTGGCCCGAGGAGGCTGCATGCAC TTTCACTGTACCCAGTGCCGCCACAGTTCTGCAGCGGCTGTACAATGCCTTTTACGCCAAG AATAAATGTCCAGAGCCTAACTGCAGGGTGAAAAAGTCCCTGCACGGCCACCACCTCGAGAC TGCTCTTCTACCTGCGGGACTGGACTGCTCTCCGGCTCAGAAGCTGCTACAGGACAATAAC GTCATGTTAATACAGAGCCTCCAGCTGGGGCCGGGCAGTCCCTGGAGGTGGCTGCCAGTG ATAGAGCAGAAGGAGGTTCCCAATGGGCTCAGGGACGAAGCTTGTGGCAAGGAACTCCAGCT GGCTATGCCGGCCTGTGCCAGGCACACTACAAAGAGTATCTTGTGAGCCTCATCAATGCCAC TCGCTGGACCCAGCCACCTTGTATGAGGTGGAAGAGCTGGAGACGGCCACTGAGCGGTACCTG CACGTACGCCCCCAGCCTTTGGCTGGAGAGGATCCCCCTGCTTACCAGGCCCGCCTTCTGCAG AAGCTGACAGAAGAGGTACCTTGGGACAGAGTATCCCCGACAGGCGGAAGTAGCTGAGGGCA AGGGTCCCGATGAGGGTCCCATGGCCTGCTCCCTCAGGAACAGCTCCAGCACCATAAAGAGG CATCTTACCACCCAGGCTTCTTGGTGGTCTTCTTCTGGTGCCACCATCTAGGGGCACCAAG GAAAGAGCGGGG		
	ORF Start: ATG at 16		ORF Stop: TAG at 3202
	SEQ ID NO: 168	1062 aa	MW at 118400.2kD
NOV43a, CG123955-01 Protein Sequence	MPGEEERAFVLVAREELASALRRDSGQAFSLEQLRPLASSLPLAARYLQLDAARLVRNAHG EPRNYLNTLSTALNILEKYGRNLLSPQRPRYWRGVKFNPNVFRSTVDVAVQGGGRDVLRLYGYT EEQPDGLSFPEGQEEPDEHQAATVTLVLLRLTELSLLQLNTHPRQQAELLEDKVEDMLQL SEFDPLLREIAPGLTTPSPVSTPGPCFLCGSAPGTLHCPSCKQALCPACDHLFHGHPRAH HLRQTLPGVLQGTHLSSLPASAQPRPQSTSLALGDSSLSPNPASAHLPWHCAACAMLEPW AVLCVACDRPRGCKGLGLGTEGPQGTGGLEPDLARGRWACQSCTFENEAAVLCSICERPLA QPPSLVVDSDAGICLQPLQGGDALLASAQSQVWYCIHCTFCNSSPGWVCMNRTSSSPIPAQ HAPRPYASSLEKGPFPKPPRRLSAPLPSSCGDPEKQRQDKMREGLQLVSMIREGEAAGACP EEIFSALQYSGTEVPLQWLRLSELPPVLEMVLAELAGQDPGLGAFSCQEARRAWLDRHGNLDEA VEECVTRRRRKVQELQSLGFGPEEGLQALFQHGDDVSRLATELQRQRLEPFRQLWDSGPEP		

TPSWDGGDKQSLVRRLAVYALPSWGRAELALSLLQETPRNYELGDVVEAVRHSQDRAFLRRL LAQECVCGWALPHNRMQALTSCECTICPDCFRQHFTIALKEKHITDMVCPACGRPDLTDDTQ LLSYFSTLDIQLRESLEPDAYALFHKKLTGVLMRDPKFLWCAQCSFGFIYEREQLEATCPQC HQTFCVRCKRQCEDFQNWKRMDPEYQAQGLAMYLOENGIDCPKCKFSYALARGGCMHFHCTQ CRHQFCSGCYNAFYAKNKCPEPNCRVKKSLLHGHHPRDCLFYLRDWTALRLQKLLQDNNVMFNT EPPAGARAVPGGGCRVIEQKEVPNGLRDEACGKETPAGYAGLCQAHYKEYLVSLINAHSLDPA TLYEVEELETATERYLHVRPQPLAGEDPPAYQARLLQKLTTEEVLGQSIPIRRRK
--

Further analysis of the NOV43a protein yielded the following properties shown in Table 43B.

Table 43B. Protein Sequence Properties NOV43a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV43a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 43C.

Table 43C. Geneseq Results for NOV43a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB50186	Human transcription factor TRFX-37 - Homo sapiens, 579 aa. [WO200172777-A2, 04-OCT-2001]	493..1062 1..579	570/579 (98%) 570/579 (98%)	0.0
ABB97407	Novel human protein SEQ ID NO: 675 - Homo sapiens, 505 aa. [WO200222660-A2, 21-MAR-2002]	630..1062 73..505	431/433 (99%) 431/433 (99%)	0.0
AAB92527	Human protein sequence SEQ ID NO:10681 - Homo sapiens, 505 aa. [EP1074617-A2, 07-FEB-2001]	630..1062 73..505	430/433 (99%) 431/433 (99%)	0.0
ABB97408	Novel human protein SEQ ID NO: 676 - Homo sapiens, 514 aa. [WO200222660-A2, 21-MAR-2002]	630..1062 73..514	432/442 (97%) 432/442 (97%)	0.0
ABB65643	Drosophila melanogaster polypeptide SEQ ID NO 23721 - Drosophila melanogaster, 2421 aa. [WO200171042-A2, 27-SEP-2001]	680..1015 2004..2381	147/378 (38%) 200/378 (52%)	7e-78

10

In a BLAST search of public sequence databases, the NOV43a protein was found to have homology to the proteins shown in the BLASTP data in Table 43D.

Table 43D. Public BLASTP Results for NOV43a				
Protein Accession Number	Protein/Organism/Length	NOV43a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96EP0	Unknown (protein for MGC:19975) - Homo sapiens (Human), 1072 aa.	1..1062 1..1072	1061/1073 (98%) 1061/1073 (98%)	0.0
Q924T7	FLJ10111 - Mus musculus (Mouse), 1057 aa.	1..1062 1..1057	918/1064 (86%) 957/1064 (89%)	0.0
Q96NF1	CDNA FLJ30980 fis, clone HHDPC2000149, highly similar to Mus musculus partial muscle protein 534 - Homo sapiens (Human), 642 aa.	430..1062 1..642	633/642 (98%) 633/642 (98%)	0.0
Q8TEI0	FLJ00217 protein - Homo sapiens (Human), 547 aa (fragment).	525..1062 1..547	538/547 (98%) 538/547 (98%)	0.0
Q96GB4	Similar to hypothetical protein FLJ10111 - Homo sapiens (Human), 539 aa.	533..1062 1..539	530/539 (98%) 530/539 (98%)	0.0

- 5 PFam analysis predicts that the NOV43a protein contains the domains shown in the Table 43E.

Table 43E. Domain Analysis of NOV43a			
Pfam Domain	NOV43a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-B_box	212..258	12/49 (24%) 29/49 (59%)	0.19
zf-RanBP	298..328	10/32 (31%) 20/32 (62%)	0.33
PHD	304..361	14/64 (22%) 40/64 (62%)	0.041
zf-RanBP	349..378	10/32 (31%) 19/32 (59%)	0.0052

Example 44.

The NOV44 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 44A.

Table 44A. NOV44 Sequence Analysis			
	SEQ ID NO: 169	2847 bp	
NOV44a, CG124672-01 DNA Sequence	ACCTTTAAGCGTCACGGGTGGGGCTGCAGCTTCTGGACCTAGGACTTTGAACATGTCGCGCCT GAAGCGGATAGCGGGCAGGATCTCCGCGCTGGTTTCAAAGCAGGTGGAAGAGACTGCGGTAC CTCGGTACCCCAAGGGCTGTTGAAGGCAGCGAGGAAGAGCGGCCAGTTAAACCTGTCGGGTAG AAACCTCAGTGAAGTGCCGCACTGTGTCTGGAGAATAAATGTGGATATCCCTGAGGAAGCTAA TCAGAATCTTTCGTTTGGTGCTACTGAAAGATGGTGGGAGCAGACAGATTGACCAAACTAAT AATATCAAACAATAAACTTCAGTCACTTACAGATGACCTGCGACTCTTGCCTGCACTGACTGT TCTTGATATACATGATAATCAGTTGACATCCCTTCCTTCTGCTATAAGAGAGCTAGAAAATCT TCAGAAACTTAATGTCAGCCATAATAAACTGAAAATACTCCCTGAAGAAATTACAAACCTAAG AAACCTGAAGTGCCTGTATCTCCAGCATAATGAATTAACCTGCATATCAGAGGGATTGAAACA ACTTTCCAATTTAGAAGATTTAGATCTTTCAAACAATCATCTTACAACCTGTTCTGTAGTTT TTCTTCTCTGTCCAGTCTGGTGGGACTCAATCTTTCTAGTAATGAACGAAGAGTTTGCCAGC AGAAATAAATAGAATGAAAAGGTTGAAGCATTGGATTGTAATTCAAATCTCTTGGAACTAT ACCTCCTGAATTGGCTGGCATGGAATCACTAGAATTGCTTTATTTGCGGAGGAATAAATTACG TTTTCTACCAGAAATTCCTTCTGTAGTCTATTGAAGGAATGCACGTAGGTGAAAACCAGAT TGAAATGTTAGAGGCAGAACATCTTAAACATCTGAATTCAATTTCTGTGCTAGACCTGAGGGA TAACAAGTTAAATCTGTTCCAGATGAAATTATACTACTACGGTCCTTGGAAAGGCTTGACCT AAGCAACAATGATATTAGTAGTCTTCCCTATTTCATTGGGGAAACCTTCATTTGAAATTTTGGC ATTAGAAGGAAATCCTTTGAGAACAATTCGAAGAGAAATTATAAGTAAAGGAACACAAGAAGT CCTAAATATCTACGAAGCAAGATCAAAGATGATGGACCTAGCCAAAGTGAGTCTGTACTGA GACTGCCATGACACTACCAAGTGAATCCAGAGTCAATATACATGCCATCATTACATTAATAAT ATTAGACTATAGTGATAAACAAGCAACTTTGATTCTGTAGGAGTGTGGATGCAAGTAAAG CAACATCGTCACTTCTATTAACCTTCAGTAAGAATCAACTATGTGAAATTCAAAAAGGATGGT AGAACTGAAGGAAATGGTTTCTGATGTCGATCTCAGTTTTAATAAACTTTCCTTTATATCCTT GGAGTTATGTGTGCTTCAGAAATGACTTTTTTAGATCTCAGGAACAATTTTTAAATCTTT GCCAGAAGAAATGGAATCACTGGTAAGACTGCAACGATCAATCTTCTTTAATAGGTTTAA AATGCTACCTGAAGTTCTATATCGTATCTTCACACTTGAACAATTCGATTAGTAATAATCA GGTTGGATCTGTGGACCTCAGAAAATGAAGATGATGGAAAATCTGACCAGTTGGACCTTCA AAATAATGACCTCTTACAAATTCACCAGAGCTCGGTAATGTGTAACTTAAGAACATTACT ATTGGATGGAATCCATTCGAGTTCCTCGAGCAGCCATATTAATGAAAGGAACAGCTGCTAT ACTTGAATATTTGAGAGACCGAATTCCTACTTAACATGGAGTTGCTTTATAACCTTGTCTATG TATTATTAACCTGGTTAATTTCTAAGGAGGATGTAACATTTGTTTTAGTATCATCTTAAAGG TGATTATTGTAATTGATCTGTAGTTTCCAGTATCACCTACCCGTTGGTATAATTAGCCTGG GCCATATTCAGTCCAGTAAATTTTTACATTTTATTAAAGATTTTGTAAAGGTGTGTGT ACATTGTGAATGGTGATAACCACAATGTGTTTATACATTTGTTCTAAATGTTTTCCTTATGAT TTATCTGCTAACTTTTCATTTTCTATAGCAAGCAGTTTTTTCAAATGAAATTTTATTATTA TGTGGTTCAGTATTATAATAACAAGCATTTTTGTAGAACTGGTTTTTTTCTCATTTATTTT TGTATTCCATACAATGTGACCAATTGACTTGAATATGACTAGCCAGTTTCTATGTTTTGTGTA GATATAAAATTAATCGAATTTTGTGAATACTGTTCTTTGGCATTAAAAAATAAGACCTTC TTATCTTGGGCCACATGTCAAAGAAAAAGGAAACAAAAATATATTAATAAATAAGACTTTTCA TTACCCATGATAGGACTTTTGTGATATGGCTAATCTCAGTACACATTTCAACTTAAACCTTT TTATTTACAGCACCATAATTTTAAATTTACTTGCAATCTGGTAAGACTAACTTGCAGTGT TTTTCTAAAAGGGAATTTGATAGGTAACTTGATTTAATAAAAAATTAATATCATTTTGTGTT ACACCAAAATATCAGAAGTAGGTTGATTAGTCATTATAACACTACCATTGATGATTCTATTAA GAAGTCAATTCAGTAGCATGTATCAATTTATATAGATAGGTAGATAGCTTTTGATGATTG AGGCATGCTTATATTGAAAAAATTGCTAATAAAGATAAATACTACATGTTTCAAGATAAA GTTACATTTTTC		
	ORF Start: ATG at 53		ORF Stop: TAA at 1859
	SEQ ID NO: 170	602 aa	MW at 68248.9kD
NOV44a, CG124672-01 Protein Sequence	MSRLKRIAGQDLRAGFKAGGRDCGTSVPQGLLKAARKSGQLNLSRNLSSEVPQCVIRINVDIP EEANQNLSFGATERWWEQTDLTKLIISNNKLQSLTDDLRLLPALTVLDIHDNQLTSLPSAIRE LENLQKLNVSNNKLKILPEEITNLRNLKCLYLQHNELTCISEGFEQLSNLEDLDSNNHLTTV PASFSSLSLVRNLSSNELKSLPAEINMRKRLKHLDCNSNLETTPELAGMESLELLYLRR NKLRLPEFPSCSLKELHVGENQIEMLEAEHLKHLNLSILVLDLRDNKLKSVPEDEIILRSLE RLDLSNNDISSLPYSGLNHLKFLALEGNPLRTIRREISKGTQEVLYLRSKIKDDGPSQSE		

	SATETAMTLPSESRVNIHAIITLKILDYSDKQATLIPDEVFDAVKSNIIVTSINFSKNQCEIP KRMVELKEMVSDVDLSFNKLSFISLELCVLQKLTFLDLRNNFLNSLPEEMESLVRLOTINLSF NRFKMLPEVLYRIFTLETILISNNQVGSVDPQMKMMENLTLDLQNNDLLQIPPELGNCVNL RTLLLDGNPFRVPRAAILMKGTAAILEYLRDRIPT		
	SEQ ID NO: 171	2712 bp	
NOV44b, CG124672-03 DNA Sequence	ACCTTTAAGCGCTCACGGGTGGGGCTGCAGCTTCTGGACCTAGGACTTTGAACATGTTCGCGCCT GAAGCGGATAGCGGGGAGGATCTCCGCGCTGGTTTCAAAGCAGGTGGAAGAGACTGCGGTAC CTCGGTACCCCAAGGGCTGTTGAAGGCAGCGAGGAAGAGCGGCCAGTTAAACCTGTTCGGGTAG AAACCTCAGTGAAGTGCCGAGTGTGTCGGAGAATAAATGTGGATATCCCTGAGGAAGCTAA TCAGAATCTTTCTGTTGGTGCTACTGAAAGATGGTGGGAGCAGACAGATTTGACCAAACATA AATATCAAACAATAAACTTCAGTCACTTACAGATGACCTGCGACTCTTGCCCTGCACTGACTGT TCTTGATATACATGATAATCAGTTGACATCCCTTCCTTCTGTATAAGAGAGCTAGAAAAATC TCAGAAACTTAATGTGAGCCATAATAAACTGAAAATACTCCCTGAAGAAATTACAAACCTAAG AAACCTGAAGTGCCTGTATCTCCAGCATAATGAATTAACCTGCATATCAGAGGGATTGAAACA ACTTTCCAATTTAGAAGATTAGATCTTTCAAACAATCATCTTACAACCTGTTCTCTGCTAGTTT TTCTTCTCTGTCCAGTCTGGTGGGACTCAATCTTTCTAGTAATGAAGTGAAGAGTTTGGCAGC AGAAATAAATAGAATGAAAAGGTTGAAGCATTTGGATTGTAATTCAAATCTCTTGAAAACAT ACCTCCTGAATTGGCTGGCATGGAATCACTAGAAATGCTTTATTGCGGAGGAATAAATTACG TTTTCTACCAGAATTTCTTCTTGTAGTCTATTGAAGGAATTGCACGTAGGTGAAAACCAGAT TGAATGTTAGAGGCAGAACATCTTAAACATCTGAATTCAAATCTTGTGCTAGACCTGAGGGA TAACAAGTTAAAAATCTGTTCCAGATGAAATATACTACTACGGTCCTTGAAAGGCTTGACCT AAGCAACAATGATATTAGTAGTCTTCCCTATTCTATTGGGGAACCTTCATTTGAAATTTTGGC ATTAGAAGGAAATCCTTTGAGAACAAATCGAAGAGAAATTATAAGTAAAGGAACACAAGAAGT CCTAAAATATCTACGAAGCAAGATCAAAGATGATGGACCTAGCCAAAGTGAGTCTGCTACTGA GACTGCCATGACACTACCAAGTGAATCCAGAGTCAATATACATGCCATCATTACATTAATAAT ATTAGACTAGTGTATAAACAAGCAACTTTGATTCTGATGAGGTGTTGATGCGAGT/ JAAAG CAACATCGTCACTTCTATTAACCTCAGTAAGAATCAACTATGTGAAATTCAAAAGGATGGT AGAAGTGAAGGAAATGGTTTCTGATGTCGATCTCAGTTTTAATAAATTTCTCTTATATCCTT GGAGTTATGTGTGCTTCAGAAATGACTTTTTTAGATCTCAGGAACAATTTTTTAAATTCCTT GCCAGAAGAAATGGAATCACTGGTAAGACTGCAAACGATCAATCTTCTCTTAATAGGTTTAA AATGCTACCTGAAGTTCTATATCGTATCTTCACACTTGAACAATCTGATTAGTAATAATCA GGTTGGATCTGTGGACCCCTCGAGCAGCCATTAATGAAAGGAACAGCTGCTATACTTGAATA TTTGAGAGACCGAATTCCTACTTAAACATGGAGTTGCTTTATAACCTTGTGATGATTATTAA CCCTGGTTAATTTAAGGAGGATGTAACATTTGTTTTAGTATCATCTTAAAGGTGATTATTG TAATTGATCTTGTAGTTTCCAGTATCACTACCCGTTGGTATAATTAGCCTGGGCCATATTC ACTGCCAGTAAATATTTTACATTTTATTTAAGATTTTTGTAAGGTTGTTGTGATCATTTGTA ATGGTGATAACCACAATGTGTTCAACATTTGTTCTAAATGTTTTGCTTATGATTTATCCTGC TAACTTTCAATTTCTTATAGCAAGCAGTTTTTCAAAAATGAATTTTATTTAATGTGGTTCA GTATTATAATAACAAAGCATTTTGTAGAACTGGTTTTTTTCTCATTTATTTTGTATTCCA TACAATGTGACCAATGACTTGAATATGACTAGCCAGTTTCTATGTTTTTGTAGATATAAAA TTAAATCGAATTTTGTGTAATACTGTTCTTTGGCATTTAAAAAATAAGACCTTCTTATCTTGG GCCACATGTCAAAGAAAAAGGAAACAAAAATATATTAATAAATAAGACTTTTCATTACCCATG ATAGGACTTTTGTGATATGGCTAATCTCAGTACACATTTCAACTTAAACCTTTTATTTTACA GCACCATAATTTTAAATTTACTTGCAATCTTGGTAAGACTAAACTTGACGTGTTTTTCTTAA AGGGAATTTGATAGGTAACTTGATTTAATAAAAATTAATATCATTTTGTGTTACACAAAA TTATCAGAAGTAGGTTGATTAGTCATTATAACACTTACCATATGATTCTATTAAGAAGTCAAT TCAGTAGCATGTATATCAATTTATATAGATAGGTAGATAGCTTTTGGATGATTGAGGCATGCT TATATTATGAAAAAATTGCTAATAAAGATAAATACTACATGTTTCAAGATAAAGTTACATTT TTC		
	ORF Start: ATG at 53		ORF Stop: TAA at 1724
	SEQ ID NO: 172	557 aa	MW at 63114.9kD
NOV44b, CG124672-03 Protein Sequence	MSRLKRIAGQDLRAGFKAGGRDCGTSVPQGLLKAARKSQGLNSGRNLSEVPQCVRINVDIP EEANQNLSFGATERWWEQTDLTKLIISNNKLQSLTDDLRLPALTVLDIHDNQLTSLPSAIRE LENLQKLVNHNKILPEEITNLRNLKCLYLQHNELTCSIEGFELSNLEDLDSNNHLLTV PASFSSLSLVRNLNSSLNELKSLPAEINRMKRLKHLDCNSNLLTIPPELAGMESLELLYLRR NKLRLPEFPSCSLLKELHVGENQIEMLEAEHLKHLNSILVLDLRDNKLSVPDEIILLRSLE RLDLSNNDISSLPYSLGNLHLKFLALEGNPLRTRREIISKGTQEVLYLRSKIKDDGPSQSE SATETAMTLPSESRVNIHAIITLKILDYSDKQATLIPDEVFDAVKSNIIVTSINFSKNQCEIP KRMVELKEMVSDVDLSFNKLSFISLELCVLQKLTFLDLRNNFLNSLPEEMESLVRLOTINLSF NRFKMLPEVLYRIFTLETILISNNQVGSVDPRAAILMKGTAAILEYLRDRIPT		
	SEQ ID NO: 173	982 bp	
NOV44c, CG124672-02 DNA	CTGCAGCTTCTGGACCTAGGACTTTGAACATGTTCGCGCCTGAAGCGGATAGCGGGCAGGATC TCCGCGCTGGTTTCAAAGCAGGTGGAAGAGACTGCGGTACCTCGGTACCCCAAGGGCTGTTGA		

Sequence	AGGCAGCGAGGAAGAGCGGCCAGTTAAACCTGTCGGGTAGAAACCTCAGTGAAGTGCCGCACT GTGTCTGGAGAATAAATGTGGATATCCCTGAGGAAGCTAATCAGAATCTTCGTTTGGTGCCTA CTGAAAGATGGTGGGAGCAGACAGATTTGACCAAATAAATATCAAAACAATAAACTTCAGT CACTTACAGATGACCTGCGACTCTTGCTGCACTGACTGTTCTTGATATACATGATAATCAGT TGACATCCCTTCCTTCTGCTATAAGAGAGCTAGAAAATCTTCAGAAACTTAATGTCAGCCATA ATAAACTGAAAATACTCCCTGAAGAAATTACAAACCTAAGAAACCTGAAGTGCTGTATCTCC AGCATAATGAATTAACCTGCATATCAGAGGGATTTGAACAACCTTCCAATTTAGAAGATTAG ATCCTTCAACAATCATCTTACAACCTGTTCTGCTAGTTTTCTTCTCTGTCCAGTCTGGTAA GACTGCAAACGATCAATCTTCTTTAATAGGTTTAAAATGCTACCCGAAGTTCTATATCGTA TCTTCACACTTGAAACAATCTGATTAGTAATAATCAGGTTGGATCTGTGGACCTCAGAAAA TGAAGATGATGGAATCTGACCACGTTGGACCTTCAAATAATGACCTTTACAAATCCAC CAGAGCTCGTAATTGTGTAACCTAAGAACATTACTACTGGATGGAATCCATTCCGAGTTT CTCGAGCAGCCATATTAATGAAAGAACAGCTGCTATACTTGAATATTTGAGAGACCGAATTC CTACTTAACATGGAGTTGCTTTATAACCTTGTCTATG		
	ORF Start: ATG at 30	ORF Stop: TAA at 951	
	SEQ ID NO: 174	307 aa	MW at 34561.3kD
NOV44c, CG124672-02 Protein Sequence	MSRLKRIAGQDLRAGFKAGGRDCGTSVPQGLLKAARKSGQLNLSGRNLSEVPQCVRINVDIP EEANQNLSFGATERWWEQDCLKLIISNNKLQSLTDDLRLLPALTVLDDIHDNQLTSLPSAIRE LENLQKLVSHNKLKILPEEITNLRNLKCLYLQHNELTCISEGFEQLSNLEDLDPSNNHLTTV PASFSSSLVRLQITNLSNFRKMLPEVLRIFTLETILISNNQVGSVDPQKMKMMENLTTL DLQNNDLLQIPPELGNCVNLRTLLLDGNPFRVPRAAILMKGTAAILLEYLRDRIPT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 44B.

Table 44B. Comparison of NOV44a against NOV44b and NOV44c.		
Protein Sequence	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV44b	1..602 1..557	525/602 (87%) 525/602 (87%)
NOV44c	1..298 1..303	225/303 (74%) 244/303 (80%)

5

Further analysis of the NOV44a protein yielded the following properties shown in Table 44C.

Table 44C. Protein Sequence Properties NOV44a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV44a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 44D.

Table 44D. Geneseq Results for NOV44a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB97457	Novel human protein SEQ ID NO: 725 - Homo sapiens, 602 aa. [WO200222660-A2, 21-MAR-2002]	1..602 1..602	602/602 (100%) 602/602 (100%)	0.0
AAB95223	Human protein sequence SEQ ID NO:17348 - Homo sapiens, 602 aa. [EP1074617-A2, 07-FEB-2001]	1..602 1..602	602/602 (100%) 602/602 (100%)	0.0
AAB73695	Human Ras-binding protein 66 - Homo sapiens, 602 aa. [WO200138367-A1, 31-MAY-2001]	1..602 1..602	602/602 (100%) 602/602 (100%)	0.0
AAU20409	Human secreted protein, Seq ID No 401 - Homo sapiens, 215 aa. [WO200155326-A2, 02-AUG-2001]	131..345 1..215	215/215 (100%) 215/215 (100%)	e-118
ABB03069	Human expressed polypeptide SEQ ID NO 42 - Homo sapiens, 215 aa. [WO200155167-A1, 02-AUG-2001]	131..345 1..215	215/215 (100%) 215/215 (100%)	e-118

In a BLAST search of public sequence databases, the NOV44a protein was found to have homology to the proteins shown in the BLASTP data in Table 44E.

5

Table 44E. Public BLASTP Results for NOV44a

Protein Accession Number	Protein/Organism/Length	NOV44a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H9A6	CDNA FLJ12889 fis, clone NT2RP2004098, weakly similar to adenylate cyclase (EC 4.6.1.1) (Unknown) (Protein for MGC:16864) - Homo sapiens (Human), 602 aa.	1..602 1..602	602/602 (100%) 602/602 (100%)	0.0
Q9HCZ4	DJ677H15.1 (A novel protein similar to leucine-rich repeat proteins) - Homo sapiens (Human), 601 aa.	1..602 1..601	601/602 (99%) 601/602 (99%)	0.0
Q9BTR7	Hypothetical 43.9 kDa protein - Homo sapiens (Human), 384 aa.	219..602 1..384	384/384 (100%) 384/384 (100%)	0.0
Q9NXC1	CDNA FLJ20331 fis, clone HEP10410 - Homo sapiens (Human), 384 aa.	219..602 1..384	383/384 (99%) 383/384 (99%)	0.0

Q9CRC8	2610040E16Rik protein (Similar to hypothetical protein FLJ20331) - Mus musculus (Mouse), 384 aa.	219..601 1..383	321/383 (83%) 355/383 (91%)	0.0
--------	--	--------------------	--------------------------------	-----

PFam analysis predicts that the NOV44a protein contains the domains shown in the Table 44F.

Table 44F. Domain Analysis of NOV44a			
Pfam Domain	NOV44a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR	106..128	9/25 (36%) 20/25 (80%)	0.0097
LRR	129..151	11/25 (44%) 20/25 (80%)	0.005
LRR	152..174	8/25 (32%) 20/25 (80%)	0.016
LRR	175..197	14/25 (56%) 22/25 (88%)	2.4e-05
LRR	198..220	10/25 (40%) 19/25 (76%)	0.013
LRR	221..243	8/25 (32%) 20/25 (80%)	0.093
LRR	244..265	9/25 (36%) 19/25 (76%)	0.36
LRR	313..335	16/25 (64%) 21/25 (84%)	0.00016
LRR	473..495	12/25 (48%) 19/25 (76%)	0.05
LRR	543..565	13/25 (52%) 20/25 (80%)	0.0012

5

Example 45.

The NOV45 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 45A.

Table 45A. NOV45 Sequence Analysis			
	SEQ ID NO: 175	2797 bp	
NOV45a, CG125900-01 DNA	ATGGCCGGCACGCGCTGGGTACTCGGGCGCTGCTCCGGGGCTGCGGCTGTAAGTGCAGCAGC TGCCGGCGCACCGGCGCCGCTGCCTGCCCTTCTACTCCGCCGCTGGCTCTATCCCGTCGGGC		

Sequence	GTCTCGGGCCGCGCCGCGCTGCTGCTGCTGCTCGGGGCGCGCGGCGCTGCCTCCAGACG CGTGGCCTCCAGACCGGGCCTGTGCTCTCCGGGAGGCTGGCGGGGCGCTCCCGCTGTGGCCACC TCTGCCGCGGGCCGCGCGCGCGCTCCTACCTGCCCTCCGTGCCTCTCTGCTGCCGAGTCG CTGGCGGCGCGCGCGCGCTCCGACGCGCAGCTACAGCCAGGAGTCCAAAACACTTACTACCTG GAAGACCTTCCACCACCCCTGAGTATGAATTGGCCCGTCCAAGTTAGAAGAGGAAGTGGAT GATGTCTTTCTCATTGAGCTCAAGGACTGCCCTGGTCATGCACTATGGAAGATGTGCTTAAC TTTTTTTCAGACTGCAGAATCCGCAACGGTGAGAATGGAATACATTTTCTCCTAAACAGAGAT GGGAAACGAAGGGGTGATGCCTTAATTGAAATGGAGTCAGAGCAGGATGTGCAGAAAGCCTTA GAGAAGCACCGCATGTACATGGGCCAGCGGTATGTGGAAGTATATGAGATAAAACAATGAAGAT GTGGATGCCTTAATGAAGAGCTTGCAGGTCAAATCTTCGCCTGTGGTAAATGATGGTGTGGTT CGTTTGAGAGGACTTCTTATAGTTGCAATGAGAAAGACATTGTAGACTTCTTTCAGGACTG AATATAGTTGACATTACTTTTGTGATGGACTATAGAGGGAGGCGAAAAACAGGGGAAGCCTAT GTGCAATTGAAGAACAGAAATGGCCAACCAAGCCCTGTTGAAACACAGGGAAGAAATGGT AATCGATACATCGAGATATTTCCAAGCAGAAGGAATGAAGTTGGAACACATGTGCGTTCTTAT AAGGGAAGAAAAATCGCATCTTTCTACTGCTAAGTATATAACTGAGCCAGAAATGGTCTTT GAAGAACATGAAGTAAATGAGGTATTTCAACCCATGACAGCTTTTGAAGTGAGAAGGAAATA GAATTGCCTAAGGAGGTGCCAGAAAAGCTTCAGAGGCTGCTGATTTTGGAACTACGCTCTTCT CTGCATTTTGTCCACATGAGAGGATTACCTTTCCAAGCCAATGCCAAGACATTATAAACTTT TTTGCTCCACTCAAGCCTGTTAGAATCACCATGGAATACAGCTCCAGTGGGAAGGCCACTGGA GAAGCTGATGTGCACTTTGAGACCCATGAGGATGCTGTTGCAGCGATGTCTAAGGATCGGTCC CACGTTTCATCATAGGTATATTGAACGTTCCTGAATTTCATGTCCAAAAGGAAAAATAAGACTCT AGGGGCTCCAGATAATAAGGGTGAAGCAAGAAGCATTTCATTGACATCTTTCTTGGACTTG GGATATACAGTTCCAGTTTATTAGCAGCAACTGCTAGGGAATGATTTTGGTGTTTTGGGTTA ATTGCTTCTAAGAAAAGTTTCATAGTGGACTGTTTAGAAGAAGAAATGAAAGATCCAGTTTGG GATTATGAAATAAACCCAAATTAATAATTTTGTTTAAACTGTCCAGGATCTGATTTAAAAAT ATGGTCTTTGTTTTATATGATTAAATGTTTGTGTTTTCATAGATGATATGTTACTCATTGTA AGCCATATTTTATTTCAGCAGTGTTCTTTAAACGGTTTCATTTAAAAAGTAACTTTTTTTT TTTGCTGTGAATTGAGTGTCTGATGTAAACTTCTCATGGAGTGAACAGTGATTTATTTT AACCAAAACATTACCAAAGCAAAGAACGGTTTCAGACCTTTGAACTGGTATGGTTTGGCAGAA TAGTTTTAAATTTTGCTGTATTTGATTACTTTAGAGATAGGAATTTTTAAAAATCAAAACAAA AATACCACAGCTTAGTGTAATGACAATTTGGCGGTTTTATGTCTTTAGAAATGTTTTGCGCTT TCTAAGCCTTGTGCTAAAGGCGTATAACGGTGGTGCCTATCTACTTAAGGGGGCATTCTAGTC TTAACCTAAAAGTTGTCTAAACTGTCCCTCCCTGGCTTTTTTTGGTTTGGGGTAGACCTAAGG GTGTTTGTGTAGTCTCAAACTGTGAAGTGACATGTGAGAACAGTCCAGACTGGTAAGAAATTT AATGGCTTCACTTGAATTTAAACAGCTCTAGATAGGAAAAAATCAGTCTCCTCATTGCTT TTTAAATGGAGTAGTACATCCCATATTTAGAACAAGTAGGGGTGCCTTGCTTAAATAAAAAAT AGCATTTAATGTATAATTGTGTGAAGGGTTTATGGATAAAGCTGTACTTCTGTGACAATGTGG CAGTACTTTCTGCTTTAATATTAACAGCTTGTTATTTAAATATTGGACAAAATGGCTGGCTT CAAAATATAGTCATTAATAAACTAACTTATGTGCACCTGTGTAGGAGAATCAAAATCCTGTA TGCTTTCTTTGCCTTGTTCTGTTCTCAGGGTGACGACTGCCACCAGGAGATGCAGTTCTAGT TCTTAAATTAATTTGCCAGGTTTCTGACAGGTGATACCTGGAAGAGAGACTATGTCTTCT CTTACTTAATACATAACCATCTTTGATTACCAGCTAAGATGCGAAATCACTGTACTGTAGTCA ATAAATGAAGACTTGTTTCAGGCTG		
	ORF Start: ATG at 1		ORF Stop: TAA at 1441
	SEQ ID NO: 176	480 aa	MW at 53153.8kD
NOV45a, CG125900-01 Protein Sequence	MAGTRWVLGALLRGCNCSSCRRTGAACLPFYSAAGSIPSGVSGRRRLLLLGAASAAASQT RGLQTGPVPPGRLAGPPAVATSAASAAAAASYPALRASLLPQSLAAAAAVPTRSYSQESKTTYL EDLPPPEYELAPSKLEEEVDVFLIRAQGLPWSCTMEDVLNFFSDCRIRNGENGIHFLNLRD GKRRGDALIEMESEQDVQKALEKHRMYMGQRVEVEYINNEDVDALMKSLQVKSSPVVNDGVV RLRGLPYSCNEKDIVDFFAGLNIVDITFVMDYRGRRTGEAYVQFEPEMANOALLKHREEIG NRYIEIFPSRRNEVRTHVGSYKGGKIASFPTAKYITEPEMVFEHEVNEVFQPMTAFESEKEI ELPKEVPEKLPEAADFGTTSSLHFVHMRGLPFQANAQDIINFFAPLKPVRITMEYSSSGKATG EADVHFETHEDAVAAMLKDRSHVHHRYIELFLNSCPKKG		

Further analysis of the NOV45a protein yielded the following properties shown in Table 45B.

Table 45B. Protein Sequence Properties NOV45a	
PSort analysis:	0.7929 probability located in mitochondrial intermembrane space; 0.7600 probability located in nucleus; 0.4699 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome)
SignalP analysis:	Cleavage site between residues 62 and 63

A search of the NOV45a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 45C.

5

Table 45C. Geneseq Results for NOV45a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB92890	Human protein sequence SEQ ID NO:11499 - Homo sapiens, 415 aa. [EP1074617-A2, 07-FEB-2001]	142..475 3..363	149/362 (41%) 215/362 (59%)	5e-71
AAU84338	Protein HNRPH2 differentially expressed in breast cancer tissue - Homo sapiens, 460 aa. [WO200210436-A2, 07-FEB-2002]	142..332 14..205	97/193 (50%) 141/193 (72%)	3e-51
ABB50269	hNRP H1 ovarian tumour marker protein, SEQ ID NO:26 - Homo sapiens, 449 aa. [WO200175177-A2, 11-OCT-2001]	142..332 3..194	99/193 (51%) 138/193 (71%)	3e-51
AAG00751	Human secreted protein, SEQ ID NO: 4832 - Homo sapiens, 308 aa. [EP1033401-A2, 06-SEP-2000]	142..332 3..194	99/193 (51%) 138/193 (71%)	3e-51
ABG02074	Novel human diagnostic protein #2065 - Homo sapiens, 479 aa. [WO200175067-A2, 11-OCT-2001]	142..332 3..194	98/193 (50%) 136/193 (69%)	2e-50

In a BLAST search of public sequence databases, the NOV45a protein was found to have homology to the proteins shown in the BLASTP data in Table 45D.

Table 45D. Public BLASTP Results for NOV45a				
Protein Accession Number	Protein/Organism/Length	NOV45a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q12849	G-rich sequence factor-I (GRSF-I) - Homo sapiens (Human), 424 aa.	1..480 1..424	413/480 (86%) 414/480 (86%)	0.0
BAC03513	CDNA FLJ33436 fis, clone BRACE2021478, highly similar to G-RICH SEQUENCE FACTOR-I - Homo sapiens (Human), 415 aa.	85..480 20..415	387/396 (97%) 389/396 (97%)	0.0
S48081	GRSF-I protein - human, 331 aa (fragment).	150..480 1..331	331/331 (100%) 331/331 (100%)	0.0
P70333	Heterogeneous nuclear ribonucleoprotein H' (HnRNP H') (FTP-3) - Mus musculus (Mouse), 449 aa.	142..332 3..194	97/193 (50%) 140/193 (72%)	7e-51
O35737	Heterogeneous nuclear ribonucleoprotein H - Mus musculus (Mouse), 449 aa.	142..332 3..194	99/193 (51%) 138/193 (71%)	7e-51

PFam analysis predicts that the NOV45a protein contains the domains shown in the Table 45E.

Table 45E. Domain Analysis of NOV45a			
Pfam Domain	NOV45a Match Region	Identities/ Similarities for the Matched Region	Expect Value
rrm	152..224	26/83 (31%) 50/83 (60%)	7.5e-05
rrm	252..321	21/77 (27%) 47/77 (61%)	1e-11
rrm	403..471	18/77 (23%) 44/77 (57%)	0.017

5

Example 46.

The NOV46 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 46A.

Table 46A. NOV46 Sequence Analysis			
	SEQ ID NO: 177	2053 bp	
NOV46a, CG126510-01 DNA Sequence	CACCTGCTGCCCACCACCCGGGAGCAGCCTTCCCTGCCCAGGCTTCAGAGTGCCCTGTGCT GCTGCCACTGCCCCCAGCTCCAGGGCCATGTCAGAGCTCCCATCTACCTCCACCAGCATG CCGCTCCTGAAGATGCCCCACCATTTCTCGGGGTGCAGCCACCCCTGCAGCGGGCACTGTGGT GGGCACTGCAGTGGGCCTCTCTCCACCCCGAGCTCTCAGCCACTCCCTAGCACTCACAGG GATCCCGGGTGCAAGGGGCACAAGTTTGACACAGTGGCCTGGCTTGCCAGCTGCCCCAGCCC TGCGAGGCAGATGAGGGGCTGGGTGAGGAAGAGGATAGCAGCTCTGAGCGAAGCTCCTGCACC TCATCCTCCACCACCAGAGAGATGGGAAGTTCTGTGACTGCTGCTACTGTGAGTTCTTCGGC		

	CACAAATGCGGAAAAGGAGAAGGCCAGTTGGCAGCAGAAGCTCTAAAGCAGGCAAAATCGTGTT TCTGGAAGCCGGGAGCCAAGGCCTGCCAGGGAGAGGCTCTTGAGTGGCCGACCGGAACTG GATCGGGTCAACAGCTTCTCTGAGCAGCCGTCTGCAGGAGATCAAAAACACTGTCAAAGACTCC ATCCGTGCCAGCTTCAGTGTGTGTGAGCTCAGCATGGACAGCAATGGCTTCTCTAAGGAGGGG GCTGCTGAGCCTGAGCCTCAGAGTCTACCCCCCTCAAACCTCAGTGGCTCCTCAGAGCAGAG CCTGACATCAACCTTGACCTGTCCCTTTGACTTTGGGCTCCCCCTCAGAACCACAGTTACAA GCTCCAGGCGAGCCAGCCCCACCATGGGCAGAAATGAGAGGCCCCCACCACCATGGACAGAG GTGAGGGGGCCCCCTCCCGGTATCGTCCCCGAGAACGGGCTCGTGAGGAGACTCAACACCGTG CCCAACCTATCCCGGGTGATCTGGGTCAAGACACCCAAGCCGGGTACCCAGCTCCGAGGAG CCAAGCTCAAAGGAAGTTCCAGTTGCAAGCAGGAGCTGCCTGAGCCTGTGTCTCAGGTGGG AAGCCACAGAAGGGCAAGAGGCAGGGCAGTCAGGCCAAGAAGAGCGAGGCAAGCCCAGCCCC CGGCCCCCAGCCAGCCTAGAGGTTCCAGTGCCAAGGGCCAGGTGCTGGCCCCAAGCAGCCA GGCAGGGTCTAGAGCTTCCAAAGTAGGCAGCTGTGCTGAGGCTGGAGAGGGGAGCCGGGG AGCCGGCCAGGACCAGGTTGGGCTGGCAGTCCCAAACTGAGAAGGAGAAGGCAGCTCCTGG CGAACTGGCCAGGCGAGGCCAAGGCACGGCTCAGGAGCAGGAGTCTGTGCAGCCCCCAGGC CCAGCAAGGCCACAGAGCTTGCCCCAGGGCAAGGGCCGAGCCGCCGAGCCGCAACAGCAG GAGAAGCCAGCCTCCTCCTTGGACGATGTGTTCTGCCAAGGACATGGACGGGGTGGAGATG GATGAGACTGACCGAGAGTGGAGTACTTTAAGAGGTTCTGTTTGGATTCTGCAAGCAGACT CGTCAGAAAGTTGCTGTGAAGTGGACCAACTTCAGCCTCAAGAAAACCACTCTAGCACAGCT CAGTGAGGCCCTGCCAGGCTGAGCTGCTTCAGGGCATCCTGAGGCCCTGACTGCCAGCTGAA GGCGTATAATTTTCCCTCCGTGTGCCCCACNTACCCGTCCAAGACCCTCTGTGCTCCCCACC ATCCTGGACCAACCAAAGCTGAACGGATGCCACACTGTGTGTTGGGGCCCCCTGACCTCAGCAG AGCCGCTTCTTGGTGCTACGCAGCCTCCACACTCAGAGCCGTGGACTGGGCTGGCCTAAGGG CCAGGGCTGATGGTACTGCTGGCCCAACACTGCTCTCTTTGTTTGGTTTTTTTGT TTTTATTTGTTTTTTTCCAATTCTTTACTTTTGATACTGTGAAGATCTTTCGTGCCGAAAGA TAAAGCAACATTTGACACAGAAAAA 		
	ORF Start: ATG at 124	ORF Stop: TGA at 1642	
	SEQ ID NO: 178	506 aa	MW at 54743.6kD
NOV46a, CG126510-01 Protein Sequence	MPLLKMPFFSGCSHPCSGHCSGPLLPSSQPLPSTHRDPGCKGHKFAHSLGACQLPQ PCEADEGLGEEEDSSSERSSCTSSSTHQRDGKFCDCYCEFFGHNAEKEKAQLAAEALKQANR VSGSREPRPARERLLEWPDRELDNRVNSFLSSRLQEIKNVKSIRASFVSCLESMDSNGFSKE GAAEPEPQSLPPSNLSSGSEQPDINLDLSPLTLGSPQNHTLQAPGEPAPPPWAEMRGPHPPWT EVRGPPPGIVPENGLVRRNLNTPNLSRVIVKTPKPGYPSSEEPSKKEVPSCQELPEPVSSG GKPQKGRQGSQAKKSEASPAPRPASLEVPASAKGOVAGPKQPGRVLELPKVGSCAEAGEGSR GSRPGPGWAGSPKTEKEKGSSWRNWPGEAKARPQEQESVQPPGPARPQSLPQKGRSRRSRNK QEKPASSLDDVFLPKMDGVEMDETREVEYFKRFCLDSAKQTRQKVAWNWTFSLKKTTPST AQ		

Further analysis of the NOV46a protein yielded the following properties shown in Table 46B.

Table 46B. Protein Sequence Properties NOV46a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV46a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 46C.

Table 46C. Geneseq Results for NOV46a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB92785	Human protein sequence SEQ ID NO:11276 - Homo sapiens, 448 aa. [EP1074617-A2, 07-FEB-2001]	110..506 52..448	397/397 (100%) 397/397 (100%)	0.0
ABB55776	Human polypeptide SEQ ID NO 158 - Homo sapiens, 586 aa. [US2001039335-A1, 08-NOV-2001]	110..506 190..586	396/397 (99%) 396/397 (99%)	0.0
AAU39067	Human secreted protein hb1041_2 - Homo sapiens, 586 aa. [WO200175068-A2, 11-OCT-2001]	110..506 190..586	396/397 (99%) 396/397 (99%)	0.0
AA Y22498	Human secreted protein sequence clone hb1041_2 - Homo sapiens, 586 aa. [WO9938959-A1, 05-AUG-1999]	110..506 190..586	396/397 (99%) 396/397 (99%)	0.0
ABG08145	Novel human diagnostic protein #8136 - Homo sapiens, 830 aa. [WO200175067-A2, 11-OCT-2001]	110..506 419..816	292/404 (72%) 312/404 (76%)	e-155

In a BLAST search of public sequence databases, the NOV46a protein was found to have homology to the proteins shown in the BLASTP data in Table 46D.

Table 46D. Public BLASTP Results for NOV46a				
Protein Accession Number	Protein/Organism/Length	NOV46a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NW00	CDNA FLJ10404 fis, clone NT2RM4000486 (Hypothetical 48.8 kDa protein) - Homo sapiens (Human), 448 aa.	110..506 52..448	397/397 (100%) 397/397 (100%)	0.0
Q96PV7	KIAA1931 protein - Homo sapiens (Human), 514 aa (fragment).	110..506 118..514	397/397 (100%) 397/397 (100%)	0.0
Q8VCA1	Similar to hypothetical protein (Hypothetical 48.1 kDa protein) - Mus musculus (Mouse), 443 aa.	110..506 52..443	325/398 (81%) 350/398 (87%)	0.0
AAH25483	Hypothetical protein - Mus musculus (Mouse), 294 aa.	110..356 52..294	198/248 (79%) 219/248 (87%)	e-106
P78311	mRNA, complete cds, clone:RES4-22A. - Homo sapiens (Human), 1224 aa.	25..501 720..1220	153/522 (29%) 236/522 (44%)	8e-38

PFam analysis predicts that the NOV46a protein contains the domains shown in the Table 46E.

Table 46E. Domain Analysis of NOV46a			
Pfam Domain	NOV46a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 47.

The NOV47 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 47A.

Table 47A. NOV47 Sequence Analysis			
	SEQ ID NO: 179	3296 bp	
NOV47a, CG127106-01 DNA Sequence	CCGCCGTTATTGTGGCCCCGACAGGCCGGGGTTACTGTGGCGACCACGAGAGCAGCTTTGGC GCTATGGAGGAGCCCGGGGCTACCCCTCAACCGTATTTGGGGCTGCTCCTGGAGGAGCTACGC AGGGTTAGGAGTCGGCTTTATGTGGGACGAGAGAAAAAGCTTGCTCTAATGCTTTCTGGACTA ATTGAAGAAAAAGTAACTACTTGAAAAATTTAGCCTTGTTCAAAAAGAGTATGAAGGCTAT GAAGTAGAGTCATCTTTAAAGGATGCCAGCTTTGAGAAGGAGGCAACAGAAGCACAAAGTTTG GAGGCAACCTGTGAAAAGCTGAACAGGTCCAATTCTGAACCTTGAGGATGAAATACTCTGTCTA GAAAAAGAGTTAAAAGAAGAGAAATCCAAACATTCTGAACAAGATGAATTGATGGCGGATATT TCAAAAAGGATACAGTCTCTAGAAGATGAGTCAAAATCCCTCAAATCACAAGTAGCTGAAGCC AAAATGACCTTCAAGATATTTCAAATGAATGAAGAACGACTGAAGATAGCAATAAAAGATGCT TTGAATGAAAATTTCTCAACTTCAGGAAAGCCAGAAACAGCTTTTGCAAGAAGCTGAAGTATGG AAAGAACAAGTGAGTGAACCTTAATAACAGAAAGTAACATTGAAGACTCCAAAGTACATGCA GAACAAGTTCTAAATGATAAAGAAAGTCACATCAAGACTCTGACTGAACGCTTGTTAAAGATG AAAGATTGGGCTGCTATGCTTGGAGAAGACATAACGGATGATGATAACTTGAATTAGAAATG AACAGTGAATCGGAAAATGGTGTCTACTTAGATAATCCTCCAAAAGGAGCTTTGAAGAAACTG ATTCATGCTGCTAAGTTAAATGCTTCTTTAAAAACCTTAGAAGGAGAAAGAAACCAATTTAT ATTCAGTTGCTGAAAGTTGATAAAACAAAGGAAGAGCTTACAGAGCATATTAATAATCTTCAG ACTGAACAAGCATCTTTCAGTCAGAAAACACACATTTTGAAAATGAGAATCAGAAGCTTCAA CAGAACTTAAAGTAATGACTGAATTATATCAAGAAAATGAAATGAACTCCACAGGAAATTA ACAGTAGAGGAAAATTTACGGTTAGAGAAAAGAGAGAACTTTCTAAAGTAGATGAAAAGATC AGCCATGCCACTGAAGAGCTGGAGACCTATAGAAAGCGAGCCAAAGATCTTGAAGAAGAATTG GAGAGAACTATTCATTCTTATCAAGGGCAGATTATTTCCCATGAGAAAAAAGCACATGATAAT TGGTTGGCAGCTCGGAATGCTGAAAGAAACCTCAATGATTTAAGGAAAGAAAATGCTCACAAC AGACAAAAATTAAGTAAACAGAGCTTAAATTTGAACCTTTAGAAAAAGATCCTTATGCACTC GATGTTCCAAATACAGCATTGGCAGAGAGCATTCCCCATATGGTCCCTCACCATTGGGTTGG CCTTCATCTGAAACAAGAGCTTTCTCTCTCCTCCAACCTTGTTGGAGGGTCCACTACACTC TCACCTTTGCTTCAGGGGGAGGAGGAAGAGGCTCACGAGGCCAGGGAATCCTTTGGACCAT CAGATTACCAATGAAAGAGGAGAATCAAGCTGTGATAGGTTAACCGATCCTCATAGGGCTCTC TCTGACACTGGGTTTCTGTCACTCCATGGGACCAGGACCGTAGGATGATGTTCTCCGCCA GGACAATCATATCCTGATTCAAGCCCTTCCTCCCAAAGGCAAGACAGATTTGTTCTAATTCT GGTAGACTGTCTGGACCAGCAGAATCAGAAGTTTAAATATGCCTTCTTTGGATAAAATGGAT GGGTCAATGCCTTCAGAAATGGAATCCAGTAGAAATGATACCAAGATGATCTTGGTAATTTA AATGTGCCTGATTCATCTCTCCTGTGAAAATGAAGCCACTGGCCCTGGCTTTGTTCTCCCA CCTCTTGCTCCAGTCAGAGGTCCATTGTTCCAGTGGATGCAAGAGGCCCATCTTGAGAAGA GGACCTCCTTTCCCCCACTCTCCAGGAGCATGTTTGGAGCTTCTCGAGATTATTTTCCA CCAGGGGATTTCCAGGTCCACCACCTGCTCCATTGCAATGAGAAATGCTATCCACCGAGG GGTCTTCTCCTTACCTTCCCAAGACCTGGATTTTCCCCCAACCCACATTCTGAAGGT AGAAGTGAGTTCCTCAGGTTGATTCCACCTTCAAATGAGCCTGCTACTGAACATCCAGAA		

	CCACAGCAAGAAACCTGACAATATTTTGGCTCTCTTCAAAAGTAATTTTGACTGATCTCATTT TCAGTTTAAGTAAGTCTGTTACTTAAGTGATTACACTTTTGCTCAAATTGAAGCTTAATGGA ATTATAATCTCAGGATAGTATTTTGTAATAAAGATGATTTAAATATGAATCTTATGAGTAA ATTATTTCAATTTTATTTTAGACGGTATAACTATTTCAATTTGATTAATCCACTATTATATAA ACAATAGTGGGAGTTTATATATGTAATCTTGCAGGTGGGGAGGCTTTAAATCTGAAGTCTG TGTCTTTATGCCAAGAACTGTATTTACTGTGGTTGTGGACAAATGTGAAAGTAACTTTATGCT TAAATAAATTATAGTTGATTAAAGATTTGTTTGGCATTGATAATAATAAAATCAGTAGTTT TCTATAACTATGGCTCTATTAATTAACCTTTTTCCTTTTACCAATAACTTTGAGGTGCAAAAC TCAAACCTTATGTGGTCTTTTGTGTTCAATTATGTTATGACAAATGTGCTCTCTTTCTGTAA ATAGACATGAGTGGCCCAAAGCAACAAATTAATACACTTTTAAAAGTCAAAATGATTATATT TTAAAGATAACCAGGATATTATCTAATGGTGAATTGTAGAATTTTGATCTTCTTATTCAGTGA GTTTCTTGACGGTTTCTTTATTGCTTTTTTCCCGCTGTTCTTTTGAAGGTATTTACTAT TTTCTGTGGAGGATATTGAGATGTACTACAGGATAACTGTAGTGAATGATGTGCATCATTTT GAGCTTTGGACTCAATATCTTTAGTGTTTCCCTAAATCAGATTTGTAGGTCATGTTAAGCTTC TTGCACATTAATATGATTATGGAAGGAAAGGCAGTGAAGCATACTAATAAACATCATAATAC TTAAAAA		
	ORF Start: ATG at 67	ORF Stop: TGA at 2347	
	SEQ ID NO: 180	760 aa	MW at 86012.4kD
NOV47a, CG127106-01 Protein Sequence	MEEPGATPQPYLGLLLEELRRVRSRLYVGREKKLALMLSGLIEEKSLLKFLSVQKEYEGYE VESSLKDASFEKEATEAQSLKLEATCEKLNRSNSELEDEILCLEKELKEEKSXHSEQDELMA KRIQSLKESKSLKQVAEAKMTFKIFQMNEERLKIAIKDALNENSQIQESQKQLQAEVVK EQVSELNKKQVTFEDSKVHAEQVLDNKESHIKTLTERLLKMKDWAAMLGEDITDDNLELEMN SESENGAYLDNPPKALKKLIHAAKLNASLKTLEGERNQIYIQLSEVDKTEELTEHIKNLQT EQASLQSENTHFENENQKLQKLVMTELYQENEMKLHRKLTVEENYRLEKEEKLKSVDEKIS HATEELETYRKRAKDLEELERTIHSYQQQIISHEKKAHDNWLAAARNLNDLRKENAHNR QKLTETELKFELLEKDPYALDVPNTAFGREHSPYGPSPPLGWPSSETRAFSLSPPTLLEGLT PLLPGGGGRGSRGPGNPLDHQITNERGESSCDRLTDPHRLSDTGFLSPPWDQDRMMFPPPG QSYPDALPPQRQDRFCNSGRLSGPAELRSFNMPSLDKMDGSMPSMESSRNDTKDDLGNLN VPDSSLPAENEATGPGFVPPPLAPVRGPLFPVDARGPFLRRGPPFPFPPPGAMFGASRDYFP GDFPGPPAPFAMRVNYPGRGFPYLPFRPGFFPPPHSEGRSEFPFGLIPPSNEPATEHPEP QQET		

Further analysis of the NOV47a protein yielded the following properties shown in Table 47B.

Table 47B. Protein Sequence Properties NOV47a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV47a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 47C.

Table 47C. Geneseq Results for NOV47a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG05280	Novel human diagnostic protein #5271 - Homo sapiens, 881 aa. [WO200175067-A2, 11-OCT-2001]	1..759 59..867	737/809 (91%) 739/809 (91%)	0.0
ABG20258	Novel human diagnostic protein #20249 - Homo sapiens, 881 aa. [WO200175067-A2, 11-OCT-2001]	1..759 59..867	732/809 (90%) 736/809 (90%)	0.0
AAY77574	Human cytoskeletal protein (HCYT) (clone 3768043) - Homo sapiens, 806 aa. [WO200006730-A2, 10-FEB-2000]	1..760 1..806	715/806 (88%) 734/806 (90%)	0.0
AAB70884	Human CTAGE-2 protein - Homo sapiens, 754 aa. [WO200127255-A2, 19-APR-2001]	15..741 29..754	632/727 (86%) 654/727 (89%)	0.0
AAM30851	Peptide #4888 encoded by probe for measuring placental gene expression - Homo sapiens, 777 aa. [WO200157272-A2, 09-AUG-2001]	1..729 1..775	616/775 (79%) 656/775 (84%)	0.0

In a BLAST search of public sequence databases, the NOV47a protein was found to have homology to the proteins shown in the BLASTP data in Table 47D.

Table 47D. Public BLASTP Results for NOV47a				
Protein Accession Number	Protein/Organism/Length	NOV47a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O15320	Meningioma-expressed antigen 6/11 (MEA6) (MEA11) - Homo sapiens (Human), 804 aa.	1..760 1..804	758/804 (94%) 759/804 (94%)	0.0
Q96SG9	BA500G10.2 (Novel protein similar to meningioma expressed antigen 6 (MEA6) and 11 (MEA11)) - Homo sapiens (Human), 825 aa (fragment).	1..760 15..816	664/804 (82%) 698/804 (86%)	0.0
Q96RT6	CTAGE-2 - Homo sapiens (Human), 754 aa.	15..741 29..754	632/727 (86%) 654/727 (89%)	0.0
AAH31065	Similar to Meningioma-expressed antigen 6/11 (MEA6) (MEA11) - Homo sapiens (Human), 745 aa.	15..730 29..743	628/716 (87%) 650/716 (90%)	0.0

O95046	WUGSC:H_DJ0988G15.3 protein (DJ1005H11.2) (WUGSC:H_DJ0988G15.3 protein) - Homo sapiens (Human), 777 aa.	1..729 1..775	612/775 (78%) 655/775 (83%)	0.0
--------	--	------------------	--------------------------------	-----

Pfam analysis predicts that the NOV47a protein contains the domains shown in the Table 47E.

Table 47E. Domain Analysis of NOV47a			
Pfam Domain	NOV47a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 48.

The NOV48 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 48A.

Table 48A. NOV48 Sequence Analysis			
	SEQ ID NO: 181	4797 bp	
NOV48a, CG127340-01 DNA Sequence	CTGGCTGGCTGGCTGTACACCTCGCTGCCGCCGGGCTTCGGGGGCACCGGGGAAGACTACAGC GAAGAGGGGATCCACTTCCCGTGGCTGCCGCCGCTCGTGTGCACCGCGCGGTATTCTTGGGC GCCGAAATCTGCCCTACTCAGTGTGTTCCGGGCACGGGTGGCCATCGCCTCGCACAGCGTG TGCTGACCCGGCTTCTGATGGCAGCCGCTTCCCGTGTGCTACCCGCTGGGCCGCTGCTG GACTGGGGCTGCGCCAGGAGATAAGCACCTTCTACACGCGGAGAGAAGTTGCTGGAGACGTTG CGGGCCGCAGACCCCTACAGTGACCTGGTGAAGGAGGAGCTCAACATCATACAGGGTGCCCTG GAGCTGCGCACCAAAGTGGTGGAGGAGGTGCTGACCCCTGGGAGACTGCTTCATGCTGCGC TCAGACGCGGTGCTCGACTTCGCCACTGTCTCCGAGATCTGCGCAGCGGTACACTCGCATC CCAGTGTACGAGGGTGACCAGCGGCACAACATTGTGGACATTTATTTGTCAAGGACTTGGCC TTCGTGGACCCGACGACTGCACCCGCTCCTCACTGTCAACCGCTTCTACAACCGGCCCTG CATTGTGTTTTCAATGACACCCGACTGGACACGGTTCTGGAGGAGTTAAGAAGGGAAAATCT CACCTGGCCATTGTCCAGCGGTGAATAATGAGGAGAGAAGGGACCTTTCTATGAGGTGATG GGCATTGTCAAGCTGGAGGATATCATAGAGGAGATTATCAAGTCGGAGATCTGGATGAAACT GATCTCTACACTGACAATCGGAAAAGCAGAGGGTCCCGAACGGGAGCGGAAGCGGCATGAC TTCTCCTTGTTTAAGCTTTCGGACACGGAGATGCGGGTGAAGATCTCACCACAGCTTCTGCTA GCCACACACCGCTTCATGGCCACAGAAGTGGAGCCCTTTAAGTCTCTGTACCTTTCGGAGAAG ATCCTGCTCCGGCTCCTGAAACATCCCAACGTGATCCAGGAGCTGAAGTTTGATGAGAAGAAC AAGAAGGCCCCGGAACACTACCTTACCAGCGCAACCGCCCTGTGGACTACTTTGTGCTGCTT CTACAGGGTAAAGTGGAGGTGGAGGTTGGTAAGGAAGGCCTTCGCTTTGAAAATGGAGCCTTT ACTTACTATGGCGTCCCAGCCATCATGACCACTGCTTGCTCAGATAATGACGTGCGGAAGGTT GGAAGTCTGGCTGGATCTTCTGTCTTTCTAAACCGGTCCCTTCTCGCTGCAGTGGGTTGAAT CGCTCTGAGTCTCCAAACCGAGAGCGCAGTGACTTTGGGGGCAGCAACACCCAGCTGTACAGC AGCAGCAACAACCTCTACATGCCTGACTACTCAGTCCACATCCTCAGCGATGTGCAGTTTGTG AAGATCACACGGCAGCAATATCAGAACGCACTCACTGCCTGCCACATGGACAGCTCACCTGTCTCT TCCCTGACATGGAGGCTTACAGACGGGACTCCACTAAGGCCCCACAACCCGGGGCACA CCCCAGACCCCTAAGGATGACCCCGCCATCACGCTCCTCAACAACAGGAACAGCCTGCCGTGC AGCCGCTCAGACGGGCTGAGAAGCCCCAGCGAGGTAGTGTACCTGAGGATGGAGGAGCTGGCC TTCACCCAGGAAGAAATGACTGACTTCGAGGAGCACAGCACACAGCAGTCAAGCTGTCTCCT GCAGCCGTTCCACGAGAGCAGCATCAGATAGTGAATGTTGTAACATCAACCTGGATACAGAG ACCAGCCCTGCAGTAGCGATTTGAGGAAAACGTGGGCAAGAAGCTGCTGAGAACCTTGAGT GGCAAAAAGGAAGAGGTCAACAGAAGGAGAGAGAACCTCTGAGGACAACCTCAATTTAAACA		

	CCTCTGATCACATGACAGGGCAAAGCCAGCATTCACTGGGTGTGTGAAATTCAGAGCTTTGG GGGAGAATCCACCCTCCCATCATCTGCTTCCCCAAGGCCTCCCAAGGTGACAGAAATGTTCT GCCTTCCCTTCCATCTCTTACACCCTAGCTGTGAGTTTGGCAGATTTTCCCTCGTTACCTCCA GTTGCACTCAGAACCTTGACATGGCCATAACAGAAGGAGGTGCCTCTGATAGAACATGCTAGA AATGGTCTTTTCCACAGCATAGTCTGGGACTGGAAAAGAGATGTCTGACTGCAAGCTGACAA GCCACTCTGGGACCCCTGATGCTCTTCTTTGTTCTTTGGGTCCCCTGATGCCATAGGAGACCT ATCGTCTTGGAACTTGCCATTCTTTCTCCAGAACAAAATGTTAACTTTCTAACACATTTTCAT GCATAGCTTGGCTCAGAAGGTGCCATTGGCAGACAGGCACATGGGAGGCTGGAGTAGGAGGT TGAAGATTAGTTTCAAGGGATGGACCAAGAATTTCCCCCAGAGCTTTAAAGAAGTGGGACTCAG CCATGTTGGCGCGTGATTGACATTACAGCACAGAAAATGTTAGTGACTGGTTTCTGTTAGA TAAGGGTTCCAGCAGCCTGGGGCAGTATGTCTCAGCTGGAATGGAAGAATGTGAGATGGAAC CTCAAGTCACTGTTTTTACCAGGGACACATCTGTTTTGGCTCCCAATCAGCAGTCTTCAATCG ATCAATAATTCTGCTCTGGAAGAGAAGGAACAGGGAGCAGAGAGACCCAATGGGAGCCAGAG ATGGAACCTTCAGGTCTTAAGTGCAATCAAAGCAAAAAACAAACAAAATCTACA'GGAAGAAC TGTAAGTGCTGAAAGCAAGTTTAGCCATGACAAACCAAGAGTGCCAGGTACGCCAAGAAAG ATACATAATCTCATGGGACTTCAGTGGGAGTTACACAGGAATGTTGAAGAATCATTCTTCTTT TTCATGCATTTGTCTTCTCCACCCCCTTACTACACCTTAGCAGATCAGCTGAGTGTACTTT ATTCCAAGAACTTACTGGATCTCTGGTTTTTCTCTGAAGTTGGGGCAGGTGCAATTCGAAG ATAACCACAGATGGCAGATGACCGGCATACCTGCTTCCAAGAATAAACAGTCTGAAAA GCAACCGCAAAGCCGGGCGCGGTGGCTCACACCTGTAATCCCAACAGTTTGGAAAGCCGAGGC GGGTGGATCACTTGAAGTCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAACCCCCATCT CTCCTGGGCATGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGGGA GGCAGAGGTTGCAGTGAGCTGAGATCACACCACTGCACTTCACTTAGGCAACAGAGCGAGAC TCTGTCTCCCCCTCAAAAAAAAAAAGAAAGAAAGAGAAAAGAAAAGAAAAGCAACCA CAGCCAGCCTTAGGGAAAACCTTGAAGTAAGTGAATTTGTCTTCAATAACTATCTCCCTT TCTGATCTGTCTCTACTCTTTAGATGTTCTCAGTCAAGTACTCACTGAACCTATTGATCGAG TGCTGTCTGTCTAAATCTCAAACCATTCCTCAACCTTCCCCCGTAGTATACCATCCAGCTTCC CTCCCCTTCTCCAAACCTCCTCCCTCCACCTCCCCACACCCATTGAGTCATTACAGGCAGG AGGGAGACTGATCATTCTCTGGGTTATCTGCATCTCAAAGAAAATGCTTACCCACAGGAAC TGTTAACTCAGGGGTTCTTAACTTGGGGTCCATCAGGGGGTCCATAGTTGGACTTCAG AGGGTCTATGAACCCCTAAACTGTGAGTATTTAATGTATTTATTCTGTATGTTTTTCCA GAGCATTAAAGCTTTCATAAGGTTCTCAAAGGTCTCAGACCTACAAAGAGTTAAACAAACAG ACAAACAAAAAAACACTACTATTTTAAATAGTGGAACTTTCAGCCAGCGTTTCTGCAATG CAGAGTGAAGTGATACTGGGCAGTTCGAGACAGGTTTTTAAATCATAAGTGGTCTTTTCAAAT GTCCATCAATTGATGGGGAAGGCTGGCACCCACCAAGAAGTGAAGTCTCTCAGAAATTCTCGG CACACCTTAGAGTATTGTACAACCAACACCCCAACATAAATTTGTCCCTCTTCCCCAACAA CCAGAGCAGGTGTTGACAGACAGGAGGCGCACAGCGTGTGGAAGTAAAGACTTTGGAGCTAGAG ATGCCTTTTCCAGCAATGATTATTGACTTCACCAACCCCTTGCCTGGCCTGGCCTGAGGCTC AGCAGTGCATGACTTCTCGTAGATACTTACAGTCATCCAGTCCCAACACCTGCTCTTGCCT GGTAGGAACAGGCGAAGTGTGAGCCCTCAATGTTGGGTACTTAGACCCAAACCAATAAATGGT GAGTTTTGAACAAGAACTACCATCATGAGGCTTCTTGCCAGCTGACCACTGGCCCCGGGT GCCTGCCTGGCTGGTCTTATCACCTGAGGCCACAGGCTCAAGCCACTGCTGTGCTATTAC CCCATCCCTTTGCAAAATCCCTATGGAGCCTGTCAACCACTCCCTCCCTATATACCCCAACC CACAAAGATTTTCTCAGGTTAAAAAAAAGTTTAAAAAAAAGATTTTAAATAAAGCATTTA TGAAGGCTTAATAAATGTAAATAATTTTAAATAAAATGAAATGCCTTCTCGAAAAAA AAAAAAA		
	ORF Start: ATG at 208		ORF Stop: TGA at 1966
	SEQ ID NO: 182	586 aa	MW at 66558.3kD
NOV48a, CG127340-01 Protein Sequence	MAAAPVVCYPLGRLLDWALRQEISTFYTREKLETLRAADPYSDLVKEELNI IQGALELRKV VEEVLTPGDCFMLRSDAVLDFATVSEILRSYTRIPVYEGDQRHNI VDLFVKDLAFVDPDD CTPLLTVTRFYNRPLHCVFNDRLDTVLEEFKKGKSHLAIQVRVNEGEGDPFYEVGMIVTLE DIEEIIKSEILDETDLTYDNRKQVRPQREKRHDFSLFKLSDEMRVKISPOLLLATHRFM ATEVEFPKSLYLSEKILLRLLKHPNVIQELKFDEKNKKAPEHYLYQRNRPVDYFVLLQGVKVE VEVGKEGLRFENGAFYYGVPAIMTTACSDNDVRKVGSLAGSSVFLNRSPSRCSGLNRSESPN RERSDFGGSNTQLYSSNNLYMPDYSVHILSDVQFVKITRQQYQNALTACHMDSSPOSPMEA FTDGDSTKAPTTRGTPQT PKDDPAITLLNNRNSLPCSRSDGLRSPSEVVYLRMEELAFTEEM TDFEEHSTQQLTSPAAPVTRAASDSECCNINLDTETSPCSSDFEENVGKLLRLTSGQKRKR SPEGERTSEDNSNLTPLIT		
	SEQ ID NO: 183	3711 bp	
NOV48b, CG127340-02 DNA Sequence	CTGGCTGGCTGGCTGTACACCTCGCTGCCGCCGGGCTTCGGGGGCACCGGGGAAGACTACAGC GAAGAGGGGATCCACTTCCCGTGGCTGCCGGCGCTCGTGTGCACCGGCGCGGTATTCTGGGC GCCGAAATCTGCCCTACTCAGTGTGTTCGCGGCACGGGCTGGCCATCGCTCTGCACAGCGTG TGCCCTGACCGGCTTCTGATGGCAGCGCTTCCCCGTGTGCTACCCGCTGGGCCGCTGCTG GACTGGGCGCTGCGCCAGGAGATAAGCACCTTCTACACGGGAGAGTGTCTGGAGACGTTG		

	CGGGCCGCAGACCCCTACAGTGACCTGGTGAAGGAGGAGCTCAACATCATACAGGGTGCCCTG GAGCTGCGCACCAGAGTTGTGGAGGAGGTGCTGGCCCCCTGGGAGACTGCTTCATGCTGCGC TCAGACGCGGTGCTCGACTTCGCCACTGTCTCCGAGATCCTGCGCAGCGGCTACACTCGCATC CCAGTGACGAGGGTGACCAGCGGCACAACATTGTGGACATTTATTGTCAAGGACTTGGCC TTCGTGGACCCCGACGACTGCACCCCGCTCCTCACTGTACC CGCTTCTACAACCGGCCCTG CATTGTGTTTTCAATGACACCCGACTGGACACGGTCTGGAGGAGTTAAGAAGGCATCAGAT AGTGAATGTTGTAACATCAACCTGGATACAGAGACCAGCCCCGTCAGTAGCGATTTTGAGGAA AACGTGGGCAAGAAGCTGCTGAGAACCTTGAGTGGCCAAAAAGGAAGGTACCAGAAGGA GAGAGAACCTCTGAGGACAACCTCAATTTAACACCTCTGATCACATGACAGGGCAAAGCCAGC ATTCACTGGGTGTGTGAAATTCAGAGCTTTGGGGGAGAATCCACCCTCCCATCATCTGCTTC CCCCAAGGCCTCCACAGGTGACAGAATGTTCTGCCTTCCCTTCCATCTCTTCAACCCTAGCT GTCAGTTTGGCAGATTTCCCTCGTTACCTCCAGTTCGACTCAGAACCTTGACATGGCCATAA CAGAAGGAGGTGCCTCTGATAGAACATGCTAGAAATGGTCTTTCCACAGCATAGCTCGGAC TGAAAAAGAGATGCTGACTGCAAGCTGACAATGCCACTCTGGGACCCCTGATGCTCTCTTT GTTCTTTGGGTCCCCTGATGCCATAGGAGACCTATCGTCTTGGAACTTGCCATTCTTTCTCTC AGAACAAAATGTTAACTTTCTAACACATTTTCATGCATAGCTTGGCTCAGAAGGTGCCATTGGC AGACAGGCACATGGGAGGCTGGAGTAGGAGGTCTGAAGATTAGTTCAGGGGATGGACCAAGAA TTTCCCCCAGAGCTTTAAAGAAGTGGGACTCAGCCATGTTGGCGCGTGATTGACATTACAGCA CAGAAAACCTGTTAGTGACTGGTTTCCTGTTAGATAAGGGTCCAGCAGCCTGGGGCAGTATGT CTCAGCTGGAATGAAAAGATGTGAGATGGAACCTCAAGTCACTGTTTTTACCAGGGACACAT CTGTTTTGGCTCCCAATCAGCAGTCTTCAATCGATCAATAATTCTGCTCTGGAAGAGAGGAA CAGGGAGCAGAGAGACCAACTGGGAGCCAGAGATGGAATTCAGGTCTTAAGTGCAATCAA AGCAAAAAACAACAAAACCTTACATGGAAAACTGTAAGTGTGAAAGCAAGTTTAGCCATGA CAAACCAAGAGTGCCAGGTGAGCAAGAAAGATACATAATCTCATGGACTTCAGTGGGAG TTACACAGGAATGTTGAAGAATCATTCTTCTTTTTCATGCATTTGCTCTCTCCACCCCTT ACTACACCCTAGCAGATCAGCTGAGTGACTTTATTCCAAGAACTTACTGGATCTCTGGTTTT TCTCTGAAGTTGGGGCAGGTGCAATTCCAAGCATAACCACAGATGGCAGAGTGACCGCGCA TACCTGCTTCCAAGAATAAAACAGTTCTGAAAAGCAACCGCAAAGCCGGCGCGGTGGCTCAC ACCTGTAATCCCAACAGTTTGGAGACCGAGGCGGTGGATCACTGAAAGTCAAGAGTTTGAG ACCAGCCTGGCCAACATGGTGAACCCCATCTCTCTGGGCATGTAGTCCAGCTACTCGGG AGGCTGAGGCAGGAGAATCGCTGAACCTGGGAGGCAGAGGTGTCAGTGAGCTGAGATCACAC CACTGCATTTCTCTAGGCAACAGAGCGAGACTCTGTCTCCCCCTCAAAAAAAGAGAA AGAAAGAAAGAGAAAAGAAAAGAAAAGCAACCACAGCCAGCCTTAGGGAAAACCTTGAAGTAA GTGAAATTTGTCTTCAGAATACTATCTCCCTTTCTGATCTGTCTCTCTCTTTAGATGTTT TCAGTCAAGTACTCACTGAACCTATTGATCGAGTGCTGTCTGCTAAATCTCCAAACCATTTCC AAACCTTTCCCGTAGTATACCATCCAGCTTCCCTCCCTTCTCTCAAAACCTCTCCCTCCAC CTCCCCACACCCATTGAGTCATTACAGGCAGGAGGGAGACTGATCATTCCTCTGGGTATCT GCATCTCAAAAGAAAATGCTTACCCACAGGAACGTTAACCTCAGGGGTTCTTAACCTGGGGTC CATCACCCAGGGGTCCATAGTTGGACTTCAGAGGGTCTATGAACCCCTAAAACTGTCAGT ATTTAATGTATTATTCTGTATGTTTTTTTCCAGAGCATTAAGCTTTCATAAGGTTCTCAAA GGTCTCAGACCTACAAAGAGTTAAAACAAACAGACAAACAAAAAACAACACTACTATTTTAA TAGTGGAACTTTCAGCCAGCGTTTCTGCAATGCAGAGTGAAGTGGATACTGGGCAGTTCGAG ACAGGTTTTAATCATAGTGGTCTTTTCAAATGTCCATCAATTGATGGGGAAGGCTGGCACC CACCAGAAGTGAAGTCTCAGAAATCTCGGCACACCCTAGAGTATTGTACAACCAACACC CCCACATAAATTTGTCCCTCTTCCCAACAACCCAGAGCAGGTGTTGCAGACAGGAGGCCA CAGCGTGTGAAGTAAAGACTTTGGAGCTAGAGATGCCTTTTCCAGCAATGATTATTGACTTC ACCACACCCCTTGCTTGCCTGGCCTGAGGCTCAGCAGTGCATGACTTCTCGTAGATAACTTC ACAGTCATCCAGTCCCAACACCTGCTCTTGCCTGGTAGGAACAGGCGAAGTGTACAGCCCTCAA TGTTGGGTACTTAGACCCAAACCAATAAATGGTGAGTTTGAACAAGAATACTACCATGTCAG GCTTCTTGGCCAGCTGACCACTGGCCCCGGGGTGCCTGCCTGGCTGGTCTTCATCACTTACGAG CCACCAGGCTCAAGCCACTGCTGTTGCATTACACCCATCCCTTTGCAAAATCCCTATGGAGCC TGTCACCACTCCCTCTATATACCCCAACCCACAAAGATTTCTTCAGGTTAAAAAAGAA GTTTAAAAAAGATTTTAAATAAAGCATTATGAAGGCTTAATAAATGTAATAAATTTT AAATAAATGAAATGCTTTTCTGGAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAAG		
	ORF Start: ATG at 208		ORF Stop: TGA at 865
	SEQ ID NO: 184	219 aa	MW at 24860.9kD
NOV48b, CG127340-02 Protein Sequence	MAAAFVPCYPLGRLLDWALRQEIISTFYTREKLETLRAADPYSDLVKEELNIIQGALELRKV VEEVLAPLGDCFMRLSDAVLDFATVSEILRSYTRIPVYEGDQRHNIVDILFVKDLAFVDPDD CTPLLTVTRFYRPLHCVFNDRLDTVLEEFKASDSECCNINLDTETSPSSDFEENVGKKL LRTLSGQKRKRSPEGERTSEDNSNLPLIT		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 48B.

Table 48B. Comparison of NOV48a against NOV48b.

Protein Sequence	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV48b	1..227 1..212	171/227 (75%) 180/227 (78%)

Further analysis of the NOV48a protein yielded the following properties shown in Table 48C.

5

Table 48C. Protein Sequence Properties NOV48a	
PSort analysis:	0.6000 probability located in nucleus; 0.4644 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1632 probability located in mitochondrial inner membrane
SignalP analysis:	Cleavage site between residues 21 and 22

A search of the NOV48a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 48D.

10

Table 48D. Geneseq Results for NOV48a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU16931	Human novel secreted protein, SEQ ID 172 - Homo sapiens, 377 aa. [WO200155441-A2, 02-AUG-2001]	211..586 2..377	375/376 (99%) 376/376 (99%)	0.0
AAB95271	Human protein sequence SEQ ID NO:17469 - Homo sapiens, 633 aa. [EP1074617-A2, 07-FEB-2001]	1..481 141..624	321/492 (65%) 376/492 (76%)	e-173
AAB95413	Human protein sequence SEQ ID NO:17804 - Homo sapiens, 853 aa. [EP1074617-A2, 07-FEB-2001]	1..481 383..844	318/487 (65%) 371/487 (75%)	e-172
AAE20847	Human gene 18 encoded secreted protein fragment, SEQ ID NO:109 - Homo sapiens, 466 aa. [WO200218435-A1, 07-MAR-2002]	1..475 1..456	288/482 (59%) 355/482 (72%)	e-157
AAE20846	Human gene 18 encoded secreted	1..475 104..559	288/482 (59%) 355/482 (72%)	e-157

	Homo sapiens, 569 aa. [WO200218435-A1, 07-MAR-2002]			
--	---	--	--	--

In a BLAST search of public sequence databases, the NOV48a protein was found to have homology to the proteins shown in the BLASTP data in Table 48E.

Table 48E. Public BLASTP Results for NOV48a				
Protein Accession Number	Protein/Organism/Length	NOV48a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NRU3	Ancient conserved domain protein 1 - Homo sapiens (Human), 586 aa.	1..586 1..586	586/586 (100%) 586/586 (100%)	0.0
Q9JIQ6	Ancient conserved domain protein 1 - Mus musculus (Mouse), 586 aa.	1..586 1..586	551/586 (94%) 569/586 (97%)	0.0
Q9NRK5	Ancient conserved domain protein 2 - Homo sapiens (Human), 633 aa.	1..481 141..624	321/492 (65%) 376/492 (76%)	e-173
Q9H952	CDNA FLJ13004 fis, clone NT2RP3000439, weakly similar to hypothetical 46.4 kDa protein in FFH-GRPE intergenic region - Homo sapiens (Human), 633 aa.	1..481 141..624	321/492 (65%) 376/492 (76%)	e-173
Q9JIM8	Ancient conserved domain protein 2 - Mus musculus (Mouse), 693 aa.	1..481 201..684	320/492 (65%) 375/492 (76%)	e-172

5

PFam analysis predicts that the NOV48a protein contains the domains shown in the Table 48F.

10

Table 48F. Domain Analysis of NOV48a			
Pfam Domain	NOV48a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 49.

The NOV49 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 49A.

Table 49A. NOV49 Sequence Analysis			
	SEQ ID NO: 185	1137 bp	
NOV49a, CG128310-01 DNA Sequence	ATGGACTCGCTGGCAGCGCCCCAGGACCGCCTGGTGGAGCAGCTGCTGTGCGCCGCGGACCCAG GCCCAGAGGCGGCTCAAGGACATTGACAAGCAGTACGTGGGCTTCGCCACACTGCCAACCCAG GTGCACCGCAAGTCGGTGAAGAAAGCCTTTGACTTCACACTCATGGTGGCTGGTGAGTCAGGC CTGGGGAAGTCCACACTGGTCCACAGCCTCTCTCTGACAGACTTGTACAAGGACCGGAAGCTG CTCAGTGCTGAGGAGCGCATCAGCCAGACGGTAGAGATTCTAAAACACACGGTGACATTGAG GAGAAGGGAGTCAAGCTGAAGCTCACCATCGTGGACACGCCGGGATTCCGGGACCGCTGTCAAC AACACCGAGTGCTGGAAGCCCATCACCAGTATGTGGACAGCAGTTTGAGCAGTACTTCCGT GATGAGAGCGCCTCAACCGAAAGAACATCCAAGACAACCGAGTGCAGTCTGCTTACTTCTC ATCTCCCCCTTCGGGCATGGGCTCGGCCAGTGGATGTGGGTTTCATGAAGGCATTGCATGAG AAGGTCAACATCGTGCTCTCATCGCCAAAGCTGACTGTCTTGTCCCCAGTGAGATCCGGAAG CTGAAGGAGCGGATCCGGGAGGAGATTGACAAGTTTGGGATCCATGTATACCAAGTTCCTTGAG TGTGACTCGGACGAGGATGAGGACTTCAAGCAGCAGGACCGGGAAGTGAAGGAGAGCGCGCCC TTCGCCGTATAGGCAGCAACACGGTGGTGGAGGCCAAGGGGAGCGGGTCCGGGCGGACTG TACCCCTGGGGGATCGTGGAGGTGGAGAACCAGGCGCATTGCGCACTTCGTGAAGCTGCGCAAC ATGCTCATCCGACGCATATGCACGACCTCAAGGACGTGACGTGCGACGTGCACTACGAGAAC TACCGCGCGCACTGCATCCAGCAGATGACCAGCAAACTGACCCAGGACCGCATGGAGAGC CCCATCCCGATCCTGCCGTGCCACCCCGGACGCCGAGACTGAGAAGCTTATCAGGATGAAG GATGAGGAAGTGAAGCGCATGCAGGAGATGCTGCAGAGGATGAAGCAGCAGATGCAGGACCAAG TGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1135
	SEQ ID NO: 186	378 aa	MW at 43844.8kD
NOV49a, CG128310-01 Protein Sequence	MDSLAAPODRLVEQLSPRTOAQRRLKIDKQYVGFATLPNQVHRKSVKKGFDFTLMVAGESG LGKSTLVHSLFSLDLYKDRKLLSAEERISQTVIELKHTVDIEEKGVKCLKTIVDTPGFGDAVN NTECWKPI TDYVDQFEQYFRDESGLNRKNIQDNRVHCCLYFISPFHGHLRPVDVGFMAKHE KVNIVPLIAKADCLVPSEIRKLKERIREEIDKFGIHVYQFPECDSDEDEDFKQDRELKESAP FAVIGSNTVVEAKQQRVGRGLYPWGI VEVENQAHCDFVKLRNMLIRTHMHLKDVTCVHYEN YRAHCIQMTSKLTQDSRMESPIPLPLPTDAETELIRMKDEELRRMQEMLQRMKQMQDQ		
	SEQ ID NO: 187	1113 bp	
NOV49b, CG128310-02 DNA Sequence	ATGAGCACAGGCCTGCGGTACAAGAGCAAGCTGGCGACCCAGAGGACAAGCAGGACATTGAC AAGCAGTACGTGGGCTTCGCCACACTGCCAACCCAGGTGCACCGCAAGTCGGTGAAGAAAGGC TTTGACTTCACACTCATGGTGGCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT CTCTTCCTGACAGACTTGTACAAGGACCGGAAGCTGCTCAGTGCTGAGGAGCGCATACGCCAG ACGGTAGAGATTCTAAAACACACGGTGGACATTGAGGAGAAGGGAGTCAAGCTGAAGCTCACC ATCGTGGACACGCCGGGATTCCGGGACGCTGTCAACAACCCAGTGTGGAAGCCCATCACC GACTATGTGGACAGCAGTTTGAGCAGTACTTCCGTGATGAGAGCGGCCTCAACCGAAAGAAC ATCCAAGACAACCGAGTGCAGTGTGCTTACTTATCTCCCTTCGGGATGGGCTGCGG CCAGTGGATGTGGGTTTCATGAAGGCATTGCATGAGAAGGTCAACATCGTGCTCTCATCGCC AAAGCTGACTGTCTTGTCCCGAGTGAAGTCCGGAAGCTGAAGGAGCGGATCCGGGAGGAGATT GACAAGTTTGGGATCCATGTATACCAAGTTCCTGAGTGTGACTCGGACGAGGATGAGGACTTC AAGCAGCAGGACCGGAACTGAAGGAGAGCGCGCCTTCGCGGTTATAGGCAGCAACACGGTG GTGGAGGCCAAGGGGACGCGGTCGGGGCCGACTGTACCCTGGGGGATCGTGGAGGTGGAG AACCAGGCGCATTGCGACTTCGTGAAGCTGCGCAACATGCTCATCCGCACGCATATGCACGAC CTCAAGGACGTGACGTGCGACGTGCACTACGAGAATAACCGCGCGCACTGCATCCAGCAGATG ACCAGCAAACTGACCCAGGACAGCCGATGGAGAGCCCCATCCCGATCCTGCCGTGCCCAAC CCGGACGCCGAGACTGAGAAGTTATCAGGATGAAGGATGAGGAAGTGAAGGCGCATGCAGGAG ATGCTGCAGAGGATGAAGCAGCAGATGCAGGACCAAGTGAAG		
	ORF Start: ATG at 1		ORF Stop: TGA at 1108
	SEQ ID NO: 188	369 aa	MW at 42776.6kD
NOV49b, CG128310-02 Protein Sequence	MSTGLRYKSLATPEDKQIDKQYVGFATLPNQVHRKSVKKGFDFTLMVAGESGLGKSTLVHS LFLTDLYKDRKLLSAEERISQTVIELKHTVDIEEKGVKCLKTIVDTPGFGDAVNNTCEWKPI DYVDQFEQYFRDESGLNRKNIQDNRVHCCLYFISPFHGHLRPVDVGFMAKHEKVNIVPLIA KADCLVPSEIRKLKERIREEIDKFGIHVYQFPECDSDEDEDFKQDRELKESAPFAVIGSNTV VEAKQQRVGRGLYPWGI VEVENQAHCDFVKLRNMLIRTHMHLKDVTCVHYENYRAHCIQMT TSKLTQDSRMESPIPLPLPTDAETELIRMKDEELRRMQEMLQRMKQMQDQ		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 49B.

Table 49B. Comparison of NOV49a against NOV49b.		
Protein Sequence	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV49b	27..378	336/352 (95%)
	18..369	337/352 (95%)

- 5 Further analysis of the NOV49a protein yielded the following properties shown in Table 49C.

Table 49C. Protein Sequence Properties NOV49a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 10 A search of the NOV49a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 49D.

Table 49D. Geneseq Results for NOV49a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB90771	Human Tumour Endothelial Marker polypeptide SEQ ID NO 275 - Homo sapiens, 369 aa. [WO200210217-A2, 07-FEB-2002]	27..378 18..369	351/352 (99%) 352/352 (99%)	0.0
AAB23259	Human cell division regulator HCDR-1 - Homo sapiens, 478 aa. [US6121019-A, 19-SEP-2000]	30..377 121..476	277/357 (77%) 312/357 (86%)	e-164
AAW73971	Human HCDR-1 protein sequence - Homo sapiens, 478 aa. [US5871973-A, 16-FEB-1999]	30..377 121..476	277/357 (77%) 312/357 (86%)	e-164

AAY23782	Human cell division regulator (HCDR) 1 - Homo sapiens, 478 aa. [US5928899-A, 27-JUL-1999]	30..377 121..476	277/357 (77%) 312/357 (86%)	e-164
AAG78669	Human bradeion protein #2 - Homo sapiens, 478 aa. [JP2001161384-A, 19- JUN-2001]	30..377 121..476	276/357 (77%) 310/357 (86%)	e-162

In a BLAST search of public sequence databases, the NOV49a protein was found to have homology to the proteins shown in the BLASTP data in Table 49E.

Table 49E. Public BLASTP Results for NOV49a				
Protein Accession Number	Protein/Organism/Length	NOV49a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BGQ3	Hypothetical 43.8 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 378 aa.	1..378 1..378	377/378 (99%) 378/378 (99%)	0.0
Q9JIM9	CDCrel-1A - Rattus norvegicus (Rat), 378 aa.	1..378 1..378	372/378 (98%) 377/378 (99%)	0.0
Q8R2F7	CDCrel-1A1 - Rattus norvegicus (Rat), 365 aa.	1..362 1..362	357/362 (98%) 360/362 (98%)	0.0
Q99648	Septin (H5) - Homo sapiens (Human), 417 aa (fragment).	27..378 66..417	351/352 (99%) 352/352 (99%)	0.0
Q99719	Septin 5 (Peanut-like protein 1) (Cell division control related protein 1) (CDCREL-1) - Homo sapiens (Human), 369 aa.	27..378 18..369	351/352 (99%) 352/352 (99%)	0.0

5

PFam analysis predicts that the NOV49a protein contains the domains shown in the Table 49F.

Table 49F. Domain Analysis of NOV49a			
Pfam Domain	NOV49a Match Region	Identities/ Similarities for the Matched Region	Expect Value
GTP_CDC	50..330	169/292 (58%) 256/292 (88%)	1.4e-185

Example 50.

The NOV50 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 50A.

Table 50A. NOV50 Sequence Analysis			
	SEQ ID NO: 189	1122 bp	
NOV50a, CG128369-01 DNA Sequence	GAGCGGGAACTGGAGCTTAAATTCTGGCGGCGAGATGGACATTCTGAAATCAGAGATCCTTC GGAAGCGGCAGCTGGTGGAGGACAGGAACCTGCTGGTGGAAAATAAAAAATATTTCAAGCGTA GTGAGCTCGCAAAAAAAGAGAGGAAGCATATTTGAAAGATGTGGCTACAAGATACAGCCAA AAGAGGAAGACACACAGAATGATCTGAAAGTTCATGAGGAAAACACCACAAATTGAAGAGTTAG AGGCGCTTGGAGAGTCCTTAGGGAAGGCGATGATCATAAGACATGGACATCATCACCAAAT TCCTGAAGTTTCTTCTTGGCGTTTGGGCTAAAGAATTGAATGCCAGAGAAGATTATGTGAAAC GCAGTGTGCAGGGTAACTGAACAGTGCAGCCAGAAACAGACCGAGTCCTACCTAAGACCAC TTTTAGAAAGCTACGAAAAGGAATCTTCTGCTGATATTAAGAATCAATAACGGATATTA TTAAATTCATGTTGCAGAGAGAATACGTGAAGGCAAATGATGCTTATCTTCAGATGGCCATTG GAAATGCGCCTTGGCCCATCGGTGTCACTATGGTTGGTATCCATGCCAGAACTGGCAGAGAAA AGATTTTTTCCAAGCATGTTGCACATGTTTTAAATGACGAAACTCAGCGGAAATATATTACAGG GATTGAAGAGGTTAATGACCATTGCCAGAAACACTTTCCTACAGACCCATCCAAATGTGTGG AGTACAATGCACCTGTGAGATCTGTGTATGGTGTGTTAATAACAATAAGAACTAGGGAAGCA GGCTGTGGACTTCTGGAATTACCAACAGGAATGAGGAAAGAAGAAAACCTGGAGTTTCCAGTCT CTGAGTTCTACCTGATGTAACCTCTTGATTGGTTTTAAGAACTTGTGTGGCCTTCATTTTCATAT CTGACTGCAAGCTGATTTTTCTTTCTTCTGCTTCATTTAATTAGTCCAAATTAAGTTTTTAAA GATTTTTCTCACAATTTAAATCCATAGACAACAGAAGGGGGTTTAAATGACCTTTTTTTCA GTTGACCCGAAAGTTGTGGTGTAGATGATTAAAAAGAAACATTGAAAAAAA		
	ORF Start: ATG at 36		ORF Stop: TGA at 771
	SEQ ID NO: 190	245 aa	MW at 28674.8kD
NOV50a, CG128369-01 Protein Sequence	MDILKSEILRRQLVEDRNLLENKKYFKRSELAKKEEYFERCGYKIQPKEEDTQNDLKVH EENTTIEELEALGESLGKGDHDKMDIITKFLKFLLGWAKELNAREDYVKRSVQGLNSATQ KQTESYLRPLFRKLRKRLNPADIKESITDIKFMQLQREYVKANDAYLQMAIGNAPWPVIGTMV GIHARTGREKIFSKHVAHVLNDETQRKYIQGLKRLMTICQKHFTDPSKCEYNAL		
	SEQ ID NO: 191	809 bp	
NOV50b, CG128369-02 DNA Sequence	GAGCGGGAACTGGAGCTTAAATTCTGGCGGCGAGATGGACATTCTGAAATCAGAGATCCTTC GGAAGCGGCAGCTGGTGGAGGACAGGAACCTGCTGGTGGAAAATAAAAAATATTTCAAGCGTA GTGAGCTCGCAAAAAAAGAGAGGAAGCATATTTGAAAGATGTGGCTACAAGATACAGCCAA AAGAGGAGGACCAGAAACCTTAACCTCATCGAATCCAGTGTAGAACTGAACTGGCAGAGG AAAAATTACCTATGACGCTTTCTAGGCAAGAGTTAGAGGCGCTTGGAGAGTCCTTAGGGAAAG GCGATGATCATAAAGACATGGACATCATCACCAAATTCCTGAAGTTTCTTCTGGCGTTTGGG CTAAGAAATTGAATGCCAGAGAAGATTATGTGAACGCGAGTGTGCAGGGTAACTGAACAGTG CGACCCAGAAACAGACCGAGTCCTACCTAAGACCCTTTTTAGAAAGCTACGGAAGGAATC TTCTGCTGATATTAAGAATCAATAACGGATATTATTAATTCATGTTGCAGAGAGAATACG TGAAGGCAAATGATGCTTATCTTCAGATGGCCATTGGAATGCGCCTTGGCCCATCGGTGTCA CTATGGTTGGTATCCATGCCAGAACTGGCAGAGAAAAGATTTTTTCCAAGCATGTTGCACATG TTTTAAATGACGAAACTCAGCGGAAATATATTACGGGATTGAAGAGGTTAATGACCATTGGC AGAAACACTTTCTCAGACCCATCCAAATGTGTGGAGTACAATGCACTGTGA		
	ORF Start: ATG at 36		ORF Stop: TGA at 807
	SEQ ID NO: 192	257 aa	MW at 29956.4kD
NOV50b, CG128369-02 Protein Sequence	MDILKSEILRRQLVEDRNLLENKKYFKRSELAKKEEYFERCGYKIQPKEEDQKPLTSSN PVLELELAEEKLPMTLSRQLEALGESLGKGDHDKMDIITKFLKFLLGWAKELNAREDYVK RSVQGLNSATQKQTESYLRPLFRKLRKRLNPADIKESITDIKFMQLQREYVKANDAYLQMAI GNAPWPVIGTMVGIHARTGREKIFSKHVAHVLNDETQRKYIQGLKRLMTICQKHFTDPSKCV EYNAL		
	SEQ ID NO: 193	843 bp	
NOV50c, CG128369-03 DNA Sequence	GAGCGGGAACTGGAGCTTAAATTCTGGCGGCGAATGGACATTCTGAAATCAGAGATCCTTCG GAAGCGGCAGCTGGTGGAGGACAGGAACCTGCTGGTGGAAAATAAAAAATATTTCAAGCGTAG TGAGCTCGCAAAAAAAGAGAGGAAGCATATTTGAAAGATGTGGCTACAAGATACAGCCAAA AGAGGAAAACACCACAATTGAAGAGTTAGAGGCGCTTGGAGAGTCCTTAGGGAAAGGCGATGA TCATAAAGACATGGACATCATCACCAAATTCCTGAAGTTTCTTCTGGCGTTTGGGCTAAAGA		

	ATTGAATGCCAGAGAAGATTATGTGAAACGCAGTGTGCAGGGTAAACTGAACAGTGCACCCAGAAACAGACCGAGTCTACCTAAGACCACTTTTAGAAAGCTACGGAAAAGGAATCTTCCTGCTGATATTAAAGAATCAATAACGGATATTATTAATTCATGTTGCAGAGAGAATACGTGAAGGCAAAATGATGCTTATCTTCAGATGGCCATTGGAATGCGCCTTGGCCCATCGGTGTCAGTGTGGTTGGTATCCATGCCAGAACTGGCAGAGAAAAGATTTTTCCAAGCATGTTGCACATGTTTTAAATGGCGAAACTCAGCGGAAATATATTCAGGGATTGAAGAGGTTAATGACCATTGTCAGAAACACTTTCCTACAGACCCATCCAAATGTGTGGAGTACAATGCACTGTGAGATCTGTGTATGGTGTGTTAATAACAATAAGAACTTAGGGAAGCAGGCTGTGGACTTCTGGAATTACCAACAGGAATGAGGAAAGAAGAAAACCTGGAAGGGCG		
	ORF Start: ATG at 35		ORF Stop: TGA at 737
	SEQ ID NO: 194	234 aa	MW at 27275.3kD
NOV50c, CG128369-03 Protein Sequence	MDILKSEILRKRQLVEDRNLLVENKKYFKRSELAKKEEYFERCGYKIQPKEENTTIEELEALGESLGKGGDDHKMDIITKFLKFLGVWAKELNAREDYVKRSVQGLNSATQKQTESYLRPLFRKLRRNLPA DIKESITDI IKFMLQREYVKANDAYLQMAIGNAPWPIGVTVVGIHARTGREKIFSKHVAHVLNGETQRKYIQGLRLMTICQKHFTDPSKVEYNAL		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 50B.

Table 50B. Comparison of NOV50a against NOV50b and NOV50c.		
Protein Sequence	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV50b	1..245	233/257 (90%)
	1..257	239/257 (92%)
NOV50c	1..245	232/245 (94%)
	1..234	233/245 (94%)

5

Further analysis of the NOV50a protein yielded the following properties shown in Table 50C.

Table 50C. Protein Sequence Properties NOV50a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3600 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV50a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 50D.

Table 50D. Geneseq Results for NOV50a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB71440	Drosophila melanogaster polypeptide SEQ ID NO 41112 - Drosophila melanogaster, 340 aa. [WO200171042-A2, 27-SEP-2001]	32..242 124..333	110/211 (52%) 148/211 (70%)	2e-59
AAG36822	Arabidopsis thaliana protein fragment SEQ ID NO: 45179 - Arabidopsis thaliana, 301 aa. [EP1033405-A2, 06-SEP-2000]	21..245 42..264	95/225 (42%) 143/225 (63%)	4e-45
AAG36821	Arabidopsis thaliana protein fragment SEQ ID NO: 45178 - Arabidopsis thaliana, 420 aa. [EP1033405-A2, 06-SEP-2000]	21..245 161..383	95/225 (42%) 143/225 (63%)	4e-45
AAG36820	Arabidopsis thaliana protein fragment SEQ ID NO: 45177 - Arabidopsis thaliana, 438 aa. [EP1033405-A2, 06-SEP-2000]	21..245 179..401	95/225 (42%) 143/225 (63%)	4e-45
AAG22300	Arabidopsis thaliana protein fragment SEQ ID NO: 25174 - Arabidopsis thaliana, 301 aa. [EP1033405-A2, 06-SEP-2000]	21..245 42..264	95/229 (41%) 145/229 (62%)	5e-45

In a BLAST search of public sequence databases, the NOV50a protein was found to have homology to the proteins shown in the BLASTP data in Table 50E.

Table 50E. Public BLASTP Results for NOV50a				
Protein Accession Number	Protein/Organism/Length	NOV50a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99633	HRP18 (Pre-mRNA splicing factor similar to S. CEREVISIAE PRP18) - Homo sapiens (Human), 342 aa.	50..245 147..342	194/196 (98%) 195/196 (98%)	e-110
Q9JKB8	Potassium channel regulatory factor - Rattus norvegicus (Rat), 342 aa.	50..245 147..342	193/196 (98%) 195/196 (99%)	e-109
Q9V437	PRP18 protein - Drosophila melanogaster (Fruit fly), 340 aa.	32..242 124..333	110/211 (52%) 148/211 (70%)	6e-59

AAM44364	POTASSIUM CHANNEL REGULATORY FACTOR - Dictyostelium discoideum (Slime mold). 389 aa.	22..240 181..388	109/224 (48%) 136/224 (60%)	2e-48
Q9SA55	F10O3.3 protein (Hypothetical 47.9 kDa protein) - Arabidopsis thaliana (Mouse-ear cress), 420 aa.	21..245 161..383	95/225 (42%) 143/225 (63%)	1e-44

PFam analysis predicts that the NOV50a protein contains the domains shown in the Table 50F.

Table 50F. Domain Analysis of NOV50a			
Pfam Domain	NOV50a Match Region	Identities/ Similarities for the Matched Region	Expect Value
hormone2	200..223	5/24 (21%) 18/24 (75%)	0.59
Prp18	91..235	102/148 (69%) 145/148 (98%)	1.2e-106

5

Example 51.

The NOV51 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 51A.

Table 51A. NOV51 Sequence Analysis			
	SEQ ID NO: 195	2468 bp	
NOV51a, CG128420-01 DNA Sequence	GTAAAGCTCTGCCATAAACTTCTAGCGTGTGCCAATGGAATTCAGCTTCGTGTTTTGCTTTC CGCTCCTCGGAACATCCGGGAGAGTTGACTTCCGGCGGCTTGTGGGAGTGCTGGTTCTGTCTCCT CCTTGGCGGTGCGGAGATGGTTGTCTTGGTTACGGGTCCTAACGGTCCCCTGCCTTGAATCC CTTGTTGAGGGCCTGCAACCTTGTGCTTCCGACTGGAGACGCCTTTGGTCCCTCGGTGTCTGC ACTGGCTGCTGGTCAAGGCTTCAGTGTGGAGTAATTGACACTTTCGAGATTGAAGAATTGGAG GAGAACTTAATGATGCACTTCACCAGAAGCAGCTACTAACATTGAGATTAGACAACCAATTG GCTTTTCAACAGAAAGATGCCAGCAAATATCAAGAATTAATGAAACAAGAAATGGAACCATT TTGTTGAGACAGAAACAAC TAGAAGAGACAAATCTTCAGCTAAGAGAAAAAGCTGGAGATGTT CGTCGAAACCTGCGTGACTTTGAGTTGACAGAAGAGCAATATATTAATANAAGCTTTTCCT GAAGATCAGCTTTCTATTCTGAATATGTATCTGTTGCTTCTATGAGCTAGTGAATCCATTA AGAAAGGAAATCTGTGAATACAAGTGAAAAAGAATATCCTAGCAGAGAATTAAGTACAAAC AAAACCAACTGAAGCAGCTGACAGAGACATATGAGGAAGATCGAAAAAACTACTCTGAAGTT CAAATTAGATGTCAACGTTTGGCCTTAGAATTAGCAGACACAAAACAGTTAATTGAGCAAGGT GACTACCGTCAAGAGAACTATGATAAAGTCAAGAGTGAACGTGATGCACTTGAACAGGAAGTA ATGAGCTTAGGAGAAAACATGAATACTTGAAGCCTCTCACATGATTCAAACAAAAGAACGA AGTGAATTATCAAAAGAGGTAGTCACCTTAGAGCAAACCTGTACTTTACTGCAAAAGGATAAA GAATATCTTAATCGCCAAAACATGGAGCTTAGTGTCTGCTGTGCTCATGAAGAGGATCGCCTT GAAAGACTTCAAGCTCAACTGGAAGAAAGCAAAAAGGCTAGAGAAGAGATGTATGAAAAATAT GTAGCATCCAGAGACCATTATAAAACAGAAATATGAAAATAAACTACATGATGAAGTGAACAA ATCAGATTGAAAAACCAAGAAATTGATCAACTTCGAAATGCCTTAGGGAAATGTATGAA CGAGAAAACAGAAATCTCCGAGAAGCAAGGATAATGCTGTGGCTGAAAAGGAACGAGCAGTG ATGGCTGAAAAGGATGCTTTAGAAAAACAGATCAGCTCTTAGACAGGTACAGAGAACTACAA		

	CTTAGTACAGAAAGCAAAGTAACAGAATTTCTCCATCAAAGTAAATTAATTAATCTTTGAAAGT GAGCGTGTTCACACTTCTGCAAGAGGAAACAGCAAGAAATCTCACACAGTGTCAATTGGAATGT GAAAAATATCAGAAAAAATTGGAGGTTTTAACCAAAGAAATTTTATAGTCTCCAAGCCTCTTCT GAAAAACGCATTACTGAACCTCAAGCACAGAAGTCAAGAGCATCAAGCAAGGCTAGACATTTAT GAGAAACTGGAAAAAGAGCTTGATGAAATAATAATGCAAAGTGCAGAAATTGAAAAATGAAGAT GAGGCTGAAAGGGTTCTTTTTCTACGGCTATGGTGCTAATGTTCCACAACAGCCAAAAAGA CGACTAAAGCAAAGTGTTCACTTGGCAAGAAGAGTGCTTCAATTAGAAAAACAAAACCTCGCTG ATTNTTAAAAGATCTGGAACATCGAAAGGACCAAGTAACACAGCTTTCAACAGGAGCTTGACA GAGGCCAATTCGCTATTAAACCAGACTCAACAGCCTTACAGGTATCTCATTGAATCAGTGCGGT CAGAGAGATTCTAAGATTGATTCACTGACGGAATCTATTGCGACAACCTTGAGAAAGATGTCAGC AACTTAAATAAAGAAAAGTCAGCTTTACTACAGACGAAGAATCAAATGGCATTAGATTTAGAA CAACTTCTAAATCATCGTGAGGAATTGGCAGCAATGAAACAGATTCTCGTTAAGATGCATAGT AAACATTCTGAGAACAGCTTACTTCTCACTAAAAACAGAACCAAAACATGTGACAGAAAAATCAG AAATCAAAGACTTTGAATGTGCCTAAAGAGCATGAAGACAATATATTTACACCTAAACCAACA CTCTTTACTAAAAAGAAGCACCTGAGTGGTCTAAGAAAAAAAAGATGAAGACCTAGTGTGTTT GGATGGGAAGCACCTGTAGACCATTATATACTCTGAAAGTTCTTTTCTGATGGAAAAACAAA TTCAGTTTAAATCGTGTAAGTCAAGCTTTTAAATAACAATGTTTATTGAACTAATATTAAT TAACAAATTCG		
	ORF Start: ATG at 418		ORF Stop: TAG at 2323
	SEQ ID NO: 196	635 aa	MW at 74956.9kD
NOV51a, CG128420-01 Protein Sequence	MKQEMETILLRQKQLEETNLQLREKAGDVRRLNLDLFELTEEQYIKLXAFPEDQLS IPEYVSVR FYELVNLPRKEICELQVKKNILAEELSTNKNQLKQLTETYEEDRKNYSEVQIRCQLALELAD TKQLIQGGDYRQENYDKVKSERDALEQEVIELRRKHEILEASHMIQTKERSELSKEVVTLEQT VTLLQKDKEYLNRQNMELSVCCAHEEDRLERLQAQLEESKKAREEMYEKYVASRDHYKTEYEN KLHDELEQIRLKTNQEIDQLRNASREMYERENRNLREARDNAVAEKERAVMAEKDALEKHDQL LDYRELQLSTESKVTEFLHQSKLSFESERVQLLQEETARNLTQCQLECEKYQKKLE VLTKE FYSLQASSEKRITELQAQNSEHQARLDIYEKLEKELDEIIMQTAEIENEDEAERVLFSYGYGA NVPTTAKRRLKQSVHLARRVLQLEKQNSLIXKRSKTSKGPSNTAFTSLTEANSLLNQTQQPY RYLIESVRQRDSKIDSLTESIAQLEKDVSNLNKEKSALLQTKNQMALDLEQLLNHREELAAMK QILVKMHSKHSNSLLLTKEPKHVTENQSKTLNVPKEHDNIFTKPTLFTKKEAPEWSKK QKMKT		

Further analysis of the NOV51a protein yielded the following properties shown in Table 51B.

Table 51B. Protein Sequence Properties NOV51a	
PSort analysis:	0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV51a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 51C.

Table 51C. Geneseq Results for NOV51a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV51a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM80122	Human protein SEQ ID NO 3768 - Homo sapiens, 690 aa. [WO200157190-A2, 09-AUG-2001]	1..559 129..689	551/561 (98%) 554/561 (98%)	0.0
AAM79138	Human protein SEQ ID NO 1800 - Homo sapiens, 612 aa. [WO200157190-A2, 09-AUG-2001]	1..473 124..596	469/473 (99%) 470/473 (99%)	0.0
AAW54095	Homo sapiens TCL52 sequence - Homo sapiens, 312 aa. [WO9812327-A2, 26-MAR-1998]	327..635 5..312	292/309 (94%) 294/309 (94%)	e-159
AAG73667	Human colon cancer antigen protein SEQ ID NO:4431 - Homo sapiens, 244 aa. [WO200122920-A2, 05-APR-2001]	315..551 1..238	209/239 (87%) 213/239 (88%)	e-105
AAG03318	Human secreted protein, SEQ ID NO: 7399 - Homo sapiens, 53 aa. [EP1033401-A2, 06-SEP-2000]	566..617 1..52	52/52 (100%) 52/52 (100%)	2e-23

In a BLAST search of public sequence databases, the NOV51a protein was found to have homology to the proteins shown in the BLASTP data in Table 51D.

Table 51D. Public BLASTP Results for NOV51a				
Protein Accession Number	Protein/Organism/Length	NOV51a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O95664	PIBF1 protein - Homo sapiens (Human), 758 aa.	1..635 124..758	633/635 (99%) 634/635 (99%)	0.0
Q8WXW3	Progesterone-induced blocking factor 1 - Homo sapiens (Human), 757 aa.	1..635 124..757	616/635 (97%) 617/635 (97%)	0.0
Q96SF4	BA555G22.2 (PIBF1 protein) - Homo sapiens (Human), 603 aa (fragment).	1..473 124..588	461/473 (97%) 462/473 (97%)	0.0
Q9CVX7	1700017E21Rik protein - Mus musculus (Mouse), 323 aa (fragment).	317..635 7..323	272/319 (85%) 288/319 (90%)	e-147
Q9D551	4930513H15Rik protein - Mus musculus (Mouse), 250 aa.	1..101 124..224	90/101 (89%) 97/101 (95%)	2e-44

Pfam analysis predicts that the NOV51a protein contains the domains shown in the Table 51E.

Table 51E. Domain Analysis of NOV51a			
Pfam Domain	NOV51a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5 Example 52.

The NOV52 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 52A.

Table 52A. NOV52 Sequence Analysis			
	SEQ ID NO: 197	1108 bp	
NOV52a, CG128519-01 DNA Sequence	GACCCGTGTTGGGATGGGAGCCCGAACCCAGGCCTGGGGCGGGCCTCAGGGACCAGCAAAT GGCCCCATCCGCTGCTCCTCAGGCCCCAGAAGCCTTCACACTCAAGGAGAAGGGGCACCTGCT GCGGCTGCCTGCGGCATTAGGAAAGCAGCTTCCAGAACTCGAGCCTGTGGGCCAGCTCAG TTCCACACAGACCAGTGATTCCACGGATGCCGCCGCTGCCAAAACCCAGTTCCTCCAGAACAT GCAGACAGCTTCAGGCGGGCCCCAGCCAGGCTCAGTGCTGTGGAGGTGGAGGCGGAGGCGGG GCGCCTGCGGAAGGCCTGCTCGCTGCTGAGACTGCGCATGAGGGAGGAGCTCTCAGCAGCCCC CATGGACTGGATGCAGGAGTACCGCTGCCTGCTCAGCTGGAGGGGCTGCAGGCCATGGTGGG CCAGTGCTGTCACAGGCTGCAGGAGCTGCGTGACGCGGTGGCGGAACAGCCACCAAGACCATG TCCTGTGGGGAGGCCCCCGGAGCCTCGCCGTCCTGTGGGGGTAGAGCGGAGCCTGCATGGAG CCCCCAGCTGCTTGTCTACTCCAGCACCCAGGAGCTGCAGACCCTGGCGGCCCTCAAGCTGCG AGTGGCTGTGCTGGACCAGCAGATCCACTTGGAAAAGTCTGATGGCTGAACCTCCTCCCCCT GGTAAGCGCTGCACAGCCGAGGGGCCGCTGCTGGCCCTGTGCCGGCTGTGCACAGCCT GCTCTGCGAGGGAGGAGCAGTGCTCCTTACCATCCTGCGGGATGAACCTGCAGTCTGAGCCTT TCCCATGCTGCCCTCGGCTGTTTCAGATGGGGATTGGGGGTGCTTCCCTGGCACTGTGCTCG GGGACCCAGAGATGCCTGTGCTTCCCTGGGAAACCTGGTGAACCTGGACCGGTGGCCTCAGT GCTCTTCTCAGGACAACTAAGCCTGCTGGTCAGGGCTGGCTTTCAGCCTTCTAAGGCTCCTG GACTCCAGAGGCCAGCGGGAGCCTTTCCTGGCTCCCTCTGTTTTCTCTACTGTAGACCAA GAGCCGCTTGTGTGATATTAAAGCCACTTAGAAAGC		
	ORF Start: ATG at 14		ORF Stop: TGA at 812
	SEQ ID NO: 198	266 aa	MW at 28644.7kD
NOV52a, CG128519-01 Protein Sequence	MGARTPRPGAGLRDQMAPSAAPQAEAFLLKEKGLLRLLPAAFRKAASQNSSLWAQLSSTQT SDSTDAAAATQFLQNMQTASGGPQPRLSAVEVEAEAGRLRKACSLRLRMRELSAAPMDWM QEYRCLLTLEGLQAMVQCLHRLQELRAAAVEQPPRPPVGRPPGASPSGGRAPWSPQLL VYSSTQELQTLAALKLRVAVLDQIHLKVLMAELLPLVSAAPQGPPLALCRAVHSLLCGE GARVLTILRDEPAV		
	SEQ ID NO: 199	831 bp	
NOV52b, CG128519-02 DNA Sequence	GACCCGTGTTGGGATGGGAGCCCGAACCCAGGCCTGGGGCGGGCCTCAGGAACAGCAAAT GGCCCCATCCGCTGCTCCTCAGGCCCCAGAAGCCTTCACACTCAAGGAGAAGGGGCACCTGCT GCGGCTGCCTGCGGCATTAGGAAAGCAGCTTCCAGAACTCGAGCCTGTGGGCCAGCTCAG TTCCACACAGACCAGTGATTCCACGGATGCCGCCGCTGCCAAAACCCAGTTCCTCCAGAACAT GCAGACAGCTTCAGGCGGGCCCCAGCCAGGCTCAGTGCTGTGGAGGTGGAGGCGGAGGCGGG GCGCCTGCGGAAGGCCTGCTCGCTGCTGAGACTGCGCATGAGGGAGGAGCTCTCGCGAGCCCC CATGGACTGGATGCAGGAGTACCGCTGCCTGCTCAGCTGGAGGGGCTGCAGGCCATGGTGGG CCAGTGCTGTCACAGGCTGCAGGAGCTGCGTGACGCGGTGGCGGAACAGCCACCAAGACCATG TCCTGTGGGGAGGCCCCCGGAGCCTCGCCGTCCTGTGGGGGTAGAGCGGAGCCTGCATGGAG CCCCCAGCTGCTTGTCTACTCCAGCACCCAGGAGCTGCAGACCCTGGCGGCCCTCAAGCTGCG		

	AGTGGCTGTGCTGGACCAGCAGATCCACTTGGAAAAGGTCCTGATGGCTGAACCTCTCCCCCT GGTAAGCGCTGCACAGCCGAGGGGCCCGCTGGCTGGCCCTGTGCCGGGCTGTGCACAGCCT GCTCTGCGAGGGAGGAGCACGTGTCTTACCATCCTGCGGGATGAACCTGCAGTCTGAGCCTT TCCCATGCTGCC		
	ORF Start: ATG at 14		ORF Stop: TGA at 812
	SEQ ID NO: 200	266 aa	MW at 28643.8kD
NOV52b, CG128519-02 Protein Sequence	MGARTPRPGAGLRNQMAPSAAPQAPAEFTLKEKGHLLRLPAAFRKAASQNSSLWAQLSSTQT SDSTDAAAATQFLQNMQTASGGPQPRLSAVEVEAEAGRLRKACSLRLRMRELSAAPMDWM QEYRCLLTLEGLQAMVGQCLHRLQELRAAVAEQPPRCPVGRPPGASPCGGRAEPWSPQLL VYSSTQELQTLAALKLRVAVLDDQIHLEKVLMAELLPLVSAAPQGGPPWLALCRAVHSLCEG GARVLTILRDEPAV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 52B.

Table 52B. Comparison of NOV52a against NOV52b.		
Protein Sequence	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV52b	1..266	212/266 (79%)
	1..266	213/266 (79%)

5

Further analysis of the NOV52a protein yielded the following properties shown in Table 52C.

Table 52C. Protein Sequence Properties NOV52a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.2413 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV52a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 52D.

Table 52D. Geneseq Results for NOV52a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB94803	Human protein sequence SEQ ID NO:15937 - Homo sapiens, 266 aa. [EP1074617-A2, 07-FEB-2001]	1..266 1..266	266/266 (100%) 266/266 (100%)	e-151
AAM50213	Human interleukin-11-like AMF7 C-terminal polypeptide - Homo sapiens, 435 aa. [WO200174897-A2, 11-OCT-2001]	1..266 170..435	266/266 (100%) 266/266 (100%)	e-151
AAM94742	Human reproductive system related antigen SEQ ID NO: 3400 - Homo sapiens, 156 aa. [WO200155320-A2, 02-AUG-2001]	33..186 1..155	134/155 (86%) 137/155 (87%)	4e-70
AAU69744	Thermus thermophilus MutS DNA mutation binding protein - Thermus thermophilus, 819 aa. [WO200173079-A2, 04-OCT-2001]	29..161 249..399	47/151 (31%) 64/151 (42%)	0.006
AA Y44931	Mammalian adipose differentiation associated protein - Mammalia, 286 aa. [WO200006591-A2, 10-FEB-2000]	14..157 84..232	41/157 (26%) 68/157 (43%)	0.33

In a BLAST search of public sequence databases, the NOV52a protein was found to have homology to the proteins shown in the BLASTP data in Table 52E.

Table 52E. Public BLASTP Results for NOV52a

Protein Accession Number	Protein/Organism/Length	NOV52a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD10336	Sequence 13 from Patent WO0174897 - Homo sapiens (Human), 435 aa (fragment).	1..266 170..435	266/266 (100%) 266/266 (100%)	e-150
Q9H872	CDNA FLJ13909 fis, clone Y79AA1000065 (Hypothetical 28.6 kDa protein) - Homo sapiens (Human), 266 aa.	1..266 1..266	266/266 (100%) 266/266 (100%)	e-150
Q9DAZ6	1600002H07Rik protein - Mus musculus (Mouse), 251 aa.	17..265 1..250	167/250 (66%) 192/250 (76%)	3e-87
Q96H61	Similar to hypothetical protein FLJ13909 - Homo sapiens (Human), 195 aa.	1..176 1..171	161/176 (91%) 161/176 (91%)	2e-84

S62790	mismatch DNA recognition protein mutS [validated] - <i>Thermus aquaticus</i> , 818 aa (fragment).	29..161 248..398	47/151 (31%) 64/151 (42%)	0.017
--------	---	---------------------	------------------------------	-------

PFam analysis predicts that the NOV52a protein contains the domains shown in the Table 52F.

Table 52F. Domain Analysis of NOV52a			
Pfam Domain	NOV52a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 53.

The NOV53 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 53A.

Table 53A. NOV53 Sequence Analysis			
	SEQ ID NO: 201	1341 bp	
NOV53a, CG128626-01 DNA Sequence	ATGTCTCAGGTGACATTTAATGATGTGGCTATAGACTTCACTCATGAAGAGTGGGGATGGCTC AGTTCTGCTCAGAGGGACTTATACAAGGATGTGATGGTCCAGAATTATGAGAACCTGGTCTCT GTAGGTCTTTCTGTAAGCCATATGTGATCACGTTATTGGAGGATGGAAAAGAGCCCTGG ATGATGGAGAAAAAAGTGTCAAAAGGTATGATTCAGATTGGGAATCAAGATGGGAAAAACAAG GAATTATCAACAAAGAAGGATAATTATGATGAAGATTCACCCCAACAGTAATAATAGAAAAA GTTGTAACAAAGTTATGAATTTCAAATTTCTAAGAAGAATTTGGAATATATAGAGAAGTTG GAAGGGAAGCATGGAAGTCAGGTAGACCATTTCAGACCAGCAATTCTCACCTCTAGAGAAAGC CCCACTGCAGACAGTGTTCACAAATACAATATATTAGAAGCACCTTCATTCAAAGTCTACT CTTTCTGAACCACAAAAAATTTCTGCTGAAGGGAATTCACACAAATATGATATATTAAGAAG AACTTACCAAAAAAGTCAGTTATAAAAAATGAGAAAGTCAATGGTGGAAAGAACTTTTGAAT TCTAATAAAAGTGGGGCAGCCTTCAGCCAGGGCAAATCTCTTACCCTTCCCCAGACTTGTAAT AGAGAGAAAATCTATACATGCAGTGAATGTGGGAAAGCCTTTGGCAAACAGTCAATCCTCAAT CGCCACTGGGAAATTCATACAGGAGAGAAGCCCTATGAATGTCGTGAATGTGGGAAGACTTTT AGCCATGGCTCATCCCTTACACGACATCTGATAAGCCATAGTGGAGAGAAACCTTACAAATGT ATTGAATGTGGGAAGGCCTTATGCCATGTCTCATCACTTACTAACCATCAGAGCACTCACACT GGAGAGAAACCATATGAATGTATGAAGTGTGGAAAGTCTTTAGTCGTGTGTCCCATCTTATT GAACATCTAAGAATTCATACTCAAGAAAACTCTATGAGTGTGATATGTGGAAAGGCCTTC ATTCATAGGTCTCTCTCATTACCATCAGAAAAATCCATACTGGAGAGAGCCCTTATGAATGT AGAGAATGTGGGAAAGCTTCTGCTGTAGCTCACACCTTACTCGACATCAAGAATTCACACT ATGGAGAAACAATATGAATGCAACAAATGTCTGAAAGTCTTTAGTAGCCTCTCATTCTTGTT CAGCATCAGAGTATTCATACTGAAGAAAAACCCTGAAGTTAGAAATGCAGGAAATCCTTCA ACCAGCTTGAATCACTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 1339
	SEQ ID NO: 202	446 aa	MW at 51273.8kD
NOV53a, CG128626-01 Protein Sequence	MSQVTFNDVAIDFTHIEWGLSSAQRDLYKDV MVQNYENLVS VGLSVTKPYVITLLEDGKEPW MMEKKLSKGMIPDWESRWENKELSTKKDNYDEDSPTVIIEKVVKQSYEFNSKKNLEYIEKL EGKHGSQVDHFRPAILTSRESPTADSVYKYNIFRSTFHSKSTLSEPKKISAEGNSHKYDILKK NLPKKSIVIKNEKVNGGKLLNSNKGAAFSQGKSLTLPQTCNREKIYTCSECGKAFGKQSILN RHWRIHTGEKPYECRECGKTFSHGSSLRHLISHSGEKPYKIECGKAFSHVSSLTNHQSTHT GEKPYECMNCGKFSRVSHLIEHLRIHTQEKLYECRICGKAFIHRSSLIHQKIHTGEKPYEC RECGKAFCCSSHLTRHQRIHTMEKQYECNKCLKVFSSLSFLVQHQSIHTEEKPLKFRNAGNPS		

TSLNH

Further analysis of the NOV53a protein yielded the following properties shown in Table 53B.

Table 53B. Protein Sequence Properties NOV53a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV53a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 53C.

Table 53C. Geneseq Results for NOV53a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB50180	Human transcription factor TRFX-31 - Homo sapiens, 570 aa. [WO200172777-A2, 04-OCT-2001]	1..445 1..448	436/448 (97%) 437/448 (97%)	0.0
AAU16110	Human novel secreted protein, Seq ID 1063 - Homo sapiens, 585 aa. [WO200155322-A2, 02-AUG-2001]	1..445 16..463	416/448 (92%) 417/448 (92%)	0.0
AAU87342	Novel central nervous system protein #252 - Homo sapiens, 398 aa. [WO200155318-A2, 02-AUG-2001]	1..380 16..395	377/380 (99%) 378/380 (99%)	0.0
ABG27051	Novel human diagnostic protein #27042 - Homo sapiens, 783 aa. [WO200175067-A2, 11-OCT-2001]	60..445 12..400	376/389 (96%) 377/389 (96%)	0.0
ABB50239	Human transcription factor TRFX-90 - Homo sapiens, 399 aa. [WO200172777-A2, 04-OCT-2001]	1..433 1..398	360/433 (83%) 375/433 (86%)	0.0

10

In a BLAST search of public sequence databases, the NOV53a protein was found to have homology to the proteins shown in the BLASTP data in Table 53D.

Table 53D. Public BLASTP Results for NOV53a				
Protein Accession Number	Protein/Organism/Length	NOV53a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9P0J4	Zinc finger protein ZNF140-like protein - Homo sapiens (Human), 411 aa.	1..446 1..411	371/446 (83%) 386/446 (86%)	0.0
Q9BZD8	ZNF140-like transcription factor (Zinc finger protein 302) - Homo sapiens (Human), 399 aa.	1..433 1..398	360/433 (83%) 375/433 (86%)	0.0
Q9NR11	ZNF135-like protein - Homo sapiens (Human), 478 aa.	1..433 1..477	336/485 (69%) 365/485 (74%)	0.0
Q96ND8	CDNA FLJ31030 fis, clone HSYRA1000137, moderately similar to zinc finger protein 184 - Homo sapiens (Human), 569 aa.	4..438 6..442	210/437 (48%) 287/437 (65%)	e-120
Q96NI8	CDNA FLJ30791 fis, clone FEBRA2000972, moderately similar to zinc finger protein 184 - Homo sapiens (Human), 536 aa.	4..438 14..448	213/437 (48%) 280/437 (63%)	e-116

PFam analysis predicts that the NOV53a protein contains the domains shown in the Table 53E.

Table 53E. Domain Analysis of NOV53a			
Pfam Domain	NOV53a Match Region	Identities/ Similarities for the Matched Region	Expect Value
KRAB	4..66	42/66 (64%) 54/66 (82%)	2.6e-36
zf-C2H2	236..258	12/24 (50%) 20/24 (83%)	9.4e-05
zf-C2H2	264..286	13/24 (54%) 20/24 (83%)	3.5e-06
zf-BED	249..287	14/52 (27%) 29/52 (56%)	0.084
zf-C2H2	292..314	13/24 (54%) 20/24 (83%)	0.00014
zf-C2H2	320..342	15/24 (62%) 18/24 (75%)	5.3e-06
zf-BED	305..343	15/52 (29%) 25/52 (48%)	0.38

zf-C2H2	348..370	10/24 (42%) 20/24 (83%)	5.2e-06
zf-C2H2	376..398	12/24 (50%) 21/24 (88%)	0.00092
zf-C2H2	404..426	10/24 (42%) 18/24 (75%)	0.00039

Example 54.

The NOV54 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 54A.

5

Table 54A. NOV54 Sequence Analysis			
	SEQ ID NO: 203	3078 bp	
NOV54a, CG128852-01 DNA Sequence	ATGTCGGCAGCCAAGGAGAACCCGTGCAGGAAATTCCAGGCCAACATCTTCAACAAGAGCAAG TGTCAAGACTGCTTCAAGCCCCGCGAGTCGCATCTGCTCAACGACGAGGACCTGACGCAGGCA AAACCCATTATGGCGGTTGGCTGCTCCTGGCTCCAGATGGGACCGACTTTGACAACCCAGTG CACCGGTCTCGGAAATGGCAGCGACGGTTCTTCATCCTTTACGAGCACGGCCTCTTGCGCTAC GCCCTGGATGAGATGCCACGACCCCTTCCTCAGGGCACCATCAACATGAACCAAGTGCACAGAT GTGGTGGATGGGAGGGCCGACGGGCCAGAAGTTCTCCCTGTGTATTCTGACGCCGTGAGAAG GAGCATTTCATCCGGGCGGAGACCAAGGAGATCGTCAGTGGGTGGCTGGAGATGCTCATGGTC TATCCCCGGACCAACAAGCAGAATCAGAAGAAGAAACGGAAGTGGAGCCCCCACACCCACAG GAGCCTGGGCTGCCAAGGTGGCTGTTACCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC ATCCCCAGTGCTGAGAAAGTCCCCACCACCAAGTCCACACTCTGGCAGGAAGAAATGAGGACC AAGGACCAGCCAGATGGCAGCAGCTGAGTCCAGCTCAGAGTCCAGCCAGAGCCAGCCTCCT GCTGCCAGTCCCTGCGGGAACCTGGGCTAGAGAGCAAAGAAGGTGAGAGCGCCATGAGTAGC GACCGCATGGAAGTGTGGCCGAAAGTCCGGGTGGAGAGCGGCTACTTCTCTCTGGAGAAGACC AAACAGGACTTGAAGGCTGAAGAACAGCAGCTGCCCGCGCGCTCTCCCTCCAGCCCCAGC ACCCCCAACCACAGGAGGTCCCAGGTGATTGAAAAGTTTGAGGCCTTGGACATTGAGAAGGCA GAGCAGATGGAGACCAATCAGTGGGGCCCTCACCATCCAGCGACACAGCCAGGGCCGCGAGC GAGAAGAGGGCGTTCCCTAGGAAGCGGACTTCACCAATGAAGCCCCCAGCTCCTCTCCCA GACGCTCGGCTTCCCCCTGTCTCCACCGAAGAGCCAAAGTCACTGAGCAGGAGGTCCACG GAGCCCTCCGTGACGCCCCGACCTGCTGAATTTCAAGAAAGGCTGGCTGACTAAGCAGTATGAG GACGGCCAGTGGAAGAAACACTGTTTGTCTCGCCGATCAAAGCCTGAGATACTACAGGGAT TCAGTGGCTGAGGAGGACGCCGACTTGGATGGAGAAATTGACTTGTCCGATGTTACCATGTC ACAGAGTATCCAGTTTCAGAGAACTATGGCTTCCAGATACATAAAAGGAGGGCGAGTTTACC CTGTGCGCATGACATCTGGGATTCGGCGGAAGTGGATCCAGACCATCATGAAGCAGCTGCAC CCGACCACTGCCCGGATGTGACCAGCTCGTTGCCAGAGGAAAAAACAAGAGCAGCTGCTCT TTTGAGACCTGCCCGAGGCTACTGAGAAGCAAGAGGCAGAGCTGGGGGAGCCGACCTGAG CAGAAGAGGAGCCGCGCACGGGAGCGGAGGCGAGAGGGCCGCTCCAAGACCTTTGACTGGGCT GAGTTCCGTCCCATCCAGCAGGCCCTGGCTCAGGAGCGGGTGGGCGCGTGGGGCCTGCTGAC ACCACAGAGCCCCTGCGCCTGAGGCGGAGCCTGGGGAGCTGGAGCGGGAGCGTGCACGGAGG CGGGAGGAGCGCGCAAGCGCTTCGGGATGCTCGACGCCACAGACGGGCCAGGCACTGAGGAT GCAGCCCTGCGCATGGAGGTGGACCGGAGCCAGGGCTGCTATGAGCGACCTCAAACCGCAT AACGTCCAGTGGAGATTGAGCAGCGGTGGCATCAGGTGGAGACCACCTCTCCGGGAAGAG AAGCAGGTGCCCATCGCGCCGCTCCACCTGTCTTCTGAAGATGGGGGTGACCGGCTCTCCACA CACGAGCTGACCTCTCTGCTCGAGAAGGAGCTGGAGCAGAGCCAGAAGGAGGCCTCAGACCTT CTGGAGCAGAACCGGCTCCTGCAGGACCAGCTGAGGGTGGCCCTGGGCGGGAGCAGAGCGCC CGTGAGGGCTACGTGCTGCAGGCCAGTGCAGCGAGGGTTTGCAGCAATGGAAGAAACGCAC CAGAAGAAGATTGAAGATCTCCAGAGGACAGCAGCGGAGCTAGAGAAGCTTCGAGAAGAG AAAGACCGCCTCTAGCCGAGGAGACAGCGGCCACCATCTCAGCCATCGAAGCCATGAAGAAG GCCACCGGGAGGAAATGGAGCGGGAGCTGGAGAAGAGCCAGCGTCCAGATCAGCAGCGTC AACTCGGATGTTGAGGCCCTGCGCGCCAGTACCTGGAGGAGCTCAGTCCGTGCAGCGGGAA CTGGAGTCTCTCGGAGCAGTACTCGCAGAAGTGCCTGGAGAATGCCATCTGGCCAGGGC CTGGAGGCGGAGCGGCGAGGCTGCGGAGTGCAGCGTGAAGACAGGAGCTCAATGCCAC AACCAGGAGCTGAACAACCGCCTGGCTGCAGAGATCACACGGTTGCGGACGCTGCTGACTGGG GACGGCGTGGGAGGCCACTGGGTACCCCTTGACAGGGCAAGGATGCCTATGAAGTAGAG		

	GTCTTATTGCGGGTAAAGGAATCGGAAATACAGTACCTGAAACAGGAGATTAGCTCCCTCAAG GATGAGCTGCAGACGGCACTGCGGGACAAGAAGTACGCAAGTACAAGTACAAAGACATCTAC ACAGAGCTCAGCATCGGAAGGCTAAGGCTGACTGTGACATCAGCAGGTGAAGGAGCAGCTC AAGGCTGCAACGGAAGCACTGGGGGAGAAGTCCCCTGACAGTCCACCGTGTCCGGATATGAT ATAATGAAATCTAAAGCAACCTGACTTCTTGAAGAAAGACAGATCCTGTGTCAACCCGGCAA CTCAGAAACATCAGGTCCAAGTCCGTAATTGAGCAGGTCTCGTGGGATACCTGA		
	ORF Start: ATG at 1		ORF Stop: TGA at 3076
	SEQ ID NO: 204	1025 aa	MW at 116459.8kD
NOV54a, CG128852-01 Protein Sequence	MSAAKENPCRKFQANIFNKSCKQCNCFKPRESHLNDEDLTQAKPIYGGWLLAPDGTDFDNPV HRSRKWQRRFFILYEHGLLRALDEMPPTLPQGTINMNQCTDVVDGEGRTGQKFSCLITPEK EHFIRAETKEIVSGWLEMLMVYPRTNKQNKKKRKEVPTTPQEPGPAKVAVTSSSSSSSSSS IPSAEKVPTTKSTLWQEMRTKDQPDGSSSLPAQSPSQSQPPAASSLREPGLESKEGESAMSS DRMDCGRKVRVESGYFSLETKQDLKAEQQLPPPLSPSPSTPNHRRSQVIEKFEALDIEKA EHMETNAVGPSPSDTRQGRSEKRAFPRKRDFNEAPPAPLPDASASPLSPHRAKSLDRRT EPSVTPDLLNFKKGWLTQYEDGQWKHWFVLADQSLRYRDSVAEEAADLDGEIDLACQYDV TEYPVQRNYGFQIHTKEGEFTLSAMTSGIRRNWIQTIMKHVHPTTAPDVTSSLPEEKNKSSCS FETCPRPTEKQEAELGEPDPEQKRSRARRRRREGRSKTFDWAEFRIQALAQERVGGVGPAD THEPLRPEAEFGELEERARRRREERRKRFGLDADTGPGTDEAALRMEVDRSPGLPMSDLKTH NVHVEIEQRWHQVETTPLEEKQVPIAPVHLSSDGGDRLSTHELTSLLEKELEQSQKASDL LEQNRLLQDLRVALGREQSAREGYVLQATCERGFAMEETHQKKIEDLQRQHRELEKLREE KDRLLAEETAATISAIEAMKNAHREEMERELEKSQRSQISSVNSDVEALRRQYLEELQSVQRE LEVLSEQYSQKCLENAHLAQALEAERQALRQCQRENQELNAHQELNRLAAEITRLRTLTTG DGGGEATGSPLAQKDAYELEVLRLVKESEIQYLKQEISSLKDELQTLALRDKKYASDKYKDIY TELSIAKAKADCISRLKEQLKAATEALGEKSPDSATVSGYDIMKSKSNPDFLKKDRSCVTRQ LRNIRSKSVIEQVSWDT		
	SEQ ID NO: 205	3990 bp	
NOV54b, CG128852-02 DNA Sequence	AGCCTGGAGCACTCTACCAGAGGGTCTCCAGCCAGCTGCAGAGCATGCACACTCTGCTGAGA GAGAAGGAGGAAGAGCTGGAGCGCATTAAAGGAAGCACATGAGAAGGTTCTGGAGAAGAAGGAG CAGGACCTCAATGAGGCTTTGGTTAAAATGGTTGCCTTGGGGAGCAGCTTAGAGGAAACAGAA ATTAAGCTCCAGGCAAAAGAAGAGATTTTAAAGGAAATTTGCAAGTGAATCTCCAAAGGACATC GAAGAGCCACGGAGTACCCCTGAAGAGAGACAGAAAGGGATGGCACTTTGCTCCAGCCACCA GTCCAAGCCACTAGGGCACCTCTAGGCCTCCACACACAAGGCTCGAGGATGAGGACGAGGAC CTGGGGGCTCCTCGGGGGAAGAGTACGGTGATGGCAGCCCCAGTAGGGAAGACAGCATGGTG CCCCAAAGTCAGTGAAGTGCTTGACAGGGAGGGCCATCAGCAGGGCACAGCCAAACTCGAC CAAGGGGCACCTGGTGTTAAAAGGCAAGAATCCGGTTCTCCACAATCCAGTGCCAAAGATAC ATTCACCCCGAAGGGTCTGAGAAGACCTGGACAGCAGCACATCTCCGACACCGCAGGAC CGGTCACCTCGGAAGAAAGCATGCTCTCAGAGCCTGCACCCAGTGTACTGCCTGCAACTGGC GACTCTGACACGTACCTCTCCATCATCCACTCCCTGGAGACCAAGCTCTACGTACAGAGGAA AAGCTCAAAGACGTGACCGTGAGGCTGGAGAGCCAGCAGGGTCAGAGCCGTGAGGCCTGCTC GCACTGCACCAACAGTGGCGGGGACCGAGGCCAGCTGCGTGAGCAGCTCCGCGCCAGCCTG CTCCAGGTTGGCGCACTGGCCTCCAGCTGGAGCAGGAGAGGCAGGAGAGGGCCAGGAGGGTT GAAGGGCATGTTGGAGAGCTTGGGACTTCCAGGTCAAGAATAGTCAGGCCCTGATGTGCCTG GAAAATGCGGAGAACAACTGAGATCTCTGCCTAGGGCCAGCCAGGAGGATGAGCAGGACGCA CGCGCAGCCTCCCTGGCAGTGTGGAGAGTGCACTCGTCAGCGCCATCCAAGCCCTGCAGCAC TGGCCGGCCCCAGCCATGGCGGGGCCGTGCACAGCTGGAGACAGGTGGCACCGAGGAGAAT GGGAAGCTGCCTCCCTGAGCAGTGTCTCCAGTCTGAGTTGACAGAGCAGGAGCAGGTGAGG CTTCTTTCTGACCAGATTGCTCTGGAGGCCCTCGTGATCAGCCAGATAGCAGATTCCCTGAAG AACACAACATCAGATGTCTCCGAATGCTCCATGAGATTTCTTGGTCAGGACAGCCACCGATG GAATCTGCTGGGGCCCCGTAGACACCTGGGCCAGGAAGGTCTAGTGAGTGGTGAGTTCTGG AGCCAGGTTGAGTCTCTGAGGAAGCACTTGGGGCACTGGGAGGAGAGGCAGTCGGTGCTCA GGAGACGGGCAGCAAAAGCATCCACAGGGCCTGGCCCCATCCTGGCCAAATGCCACATGGGTG AGGGCAGAGCTCAGCTTTGCCACACAGTCAGTGAGGGAGTCGTTCCACCCGAGGCTACAGAGC ATCCAGGAGACCTGCGGGGCACCCAGAGCGCCCTGCGGCAGCACAAATGCTGTGAGGGAA ATCCTGGGAGCCTACCAAACCCAGACTTTGAAAGAGTGATGCAGCAGGTTCTGGAAGCCCTC AGGCTTCCAGCGGGCCATGAAGATGGTGTTCAGCTGTCTGGGACCTGAGCCCTTAGGAGAA GTCCTGGGCGGAGACTCAGACAGCTCTCAGGAGCCCTCGATGTGTCTGACCAGAGCCCTGGG GCCTTTGTTGCTATTAGGAGGAGCTTGCCAGCAGCTGAAGGAGAAGGCCAGCCTCTTAGAG GAGATAGCGGCTGCCTTACCATCTCTGCCACCTGTGGAATCGCTGAGAGATTGCCAGAAGCTT CTCCAGGTGTCCAGAGTCTCTCGTATAACACTTGTGTTGGGAGGCCTCGGTCACTATTCTTCA TTGTTGGTTCAAGATGCCATTATTAGGCCCCAGGTTTGCTATGCGTCTCGCAGAAATCCGGCTA GAATATGAGAAGGAGCTCCAGCTCTGAAGGAGTCTGGCAAACCCGGGAGCCCTCTGTCTCA GAGCAGGCACAGGCAGCCGGGCTGAGGGAGGAGTATGAGGAGCTTCTCCGCAAGCAGAAAG AGCGAGTACCTGGATGTGATCGCATTTGTTGAAAGGGAGAATGCAGAGCTCAAGGCCAAGGCC GCCAGCTAGACCATCAGCAGAGTGTCTGGAGGATGCAGAGAGCAAGCACAGCATGAGCATG		

	TTCACCTGCGGGGAGGTATGAGGAGGAGATTCCGGTGTGTGGTGGAGAGCTGACCAGGACC GAGAGCACACTGCAGGCTGAGCGCAGCCGGTCTCTGAGCCAGCTGGATGCCCTCGGTGAGAGAC AGGCAGGACATGGAGAGGCATCATGGTGAGCAGATACAGACCCTGGAGGACAGGTTCCAGCTC AAGGTCCGGGAGCTGCAGACGATCCACGAGGAGGAGCTGAGGACCCTGAGGAGGAGCTACTCG CAGAGCCTGAGGTGCCCTTCAGGACACCTCTGCCTCCACCAGGGGCCACACCCCAAGGCCCTG CCAGCCCCCTGCCCCAACTGGCAGGCCACCCAGGGAGAGGCTGACTCCATGACGGGGCTGAGG GAGCGCATCCAGGAGCTGGAGGCCAGATGGATGTCTATGCGGGAGGAGCTGGGACACAAGGAC CTGGAGGGGCGACGCGGCCACACTGCGTGAGAAGTACCAGAGGGACTTGAGAGGCCCTTAAGGCC ACGTGCGAGCGAGGGTTTGAGCAATGGAAGAAACGCACCAGAAGAAGATTGAAGATCTCCAG AGGCAGCACCAGCGGGAGCTAGAGAACTTCGAGAAGAGAAAGACCGCCTCCTAGCCGAGGAG ACAGCGGCCACCATCTCAGCCATCGAAGCCATGAAGAACGCCACCAGGGAGGAAATGGAGCGG GAGCTGGAGAAGAGCCAGCGGTCCAGATCAGCAGCGTCAACTCGGATGTTGAGGCCCTGCGG CGCCAGTACCTGGAGGAGCTGCAGTCCGTGACGCGGGAAGTGGAGGTCTCTCGGAGCAGTAC TCGCAAGTGCCTGGAGAATGCCATCTGGCCAGGCGCTGGAGGCCGAGCGGAGGCCCTG CGGCAGTGCCAGCGTGAGAACCAGGAGCTCAATGCCACAACCAGGAGCTGAACAACCGCCTG GCTGCAGAGATCACACGGTTGCGGACGCTGCTGACTGGGACGCGCGTGGGAGGCCACTGGG TCACCCCTTGACAGGGCAAGGATGCCTATGAAGTAGAGGTCTTATTGCGGGTAAAGGAATCG GAAATACAGTACCTGAAACAGGAGATTAGTCCCTCAAGGATGAGCTGCAGACGGCACTGCGG GACAAGAAGTACGCAAGTGACAAGTACAAAGACATCTACACAGAGCTCAGCATCGCGAAGGCT AAGGCTGACTGTGACATCAGCAGGTTGAAGGAGCAGCTCAAGGCTGCAACGGAAGCACTGGGG GAGAAGTCCCTGACAGTGCCACGGTGTCCGATATGATATAATGAAATCTAAAAGCAACCCCT GACTTCTTGAAGAAAGACAGATCCTGTGTCAACCGGCAACTCAGAAACATCAGGTCCAAGAGT CTGAAGGAAGGCCTGACGGTGCAAGAACGGTTGAAGCTCTTTGAATCCAGGAGCTGAAGAAA GACTAGGTGTGTCCTATCCAAGTTGAGCACGCGCCTTCCCCAGCTTGAGCAGCACACCCCAA GCGCTGCTTTTACCTGTACCTTTGTTTATTATTATTATTATTGCTGTTGTTGTCATCG TTAACTGTGGGCATGGAATGC		
	ORF Start: ATG at 46		ORF Stop: TAG at 3847
	SEQ ID NO: 206	1267 aa	MW at 142689.7kD
NOV54b, CG128852-02 Protein Sequence	MHTLLREKEEELERIKEAHEKVLEKKEQDLNEALVKMVALGSSLEETEIKLQAKEEILRFAS ESPKDMEEPRSTPEETERDGTLLPGQPVAQTRAPLGLPHTRLEDEDEDLGAPPGEYVGDGSPS REDMSVPPKSVVEVLDREGHQGTAKLDQAGPVKQRIRFSTIQCRQYIHPEGSEKTWTSSTS SDTSQDRSPSEESMSSEPAFSLPATGDSDTYLSIHSLETCLYVTEELKLDVTVRLESQQG SREALLALHHQWAGTEAQLREQLRASLLQVGALASQLEQERQERARRVEGHVGLGDFQVKN QALMCLENCREQLRSLPRASQEDQDARAASLASVESALVSAIQALQHWPAHGGARAQLET GGTEENGKPAQLQCSQSELTEQEQVRLSDQIALEASLSIQIADSLKNTTSDVSRMLHEISW SGQPPMESAGAPVDTWARKVLVDGEFWSQVESLRKHLGTLGGEAVGASGDGQSQIPQGLAPIL ANATWVRAELSFATQSVRESFHRRLQSIQETLRGTQTALRQHKCLLREILGAYOTPDFERVMO QVLEALRLPAGHEDGVQLSWDLSPLGEVLGRSDSSQEPFDVSDQSPGAFVAIQEELAQQKE KASLLEEIAAALPSLPPVESLRDCQKLLQVSQSLSYNTCLGGLGQYSSLLVQDAIIQAQVCYA SCRIRLEYEKELQCKESWQTPRESCSEQAQAARALREYEELLRKQKSEYLDVIAIVERENA ELKAKAAQLDHQQQCLEDASKHSMSMFTLRGRYEEIIRCVEQLTRSTLQAERSRVLSQL DASVRDRQDMERHHGEQIQTLDRFQLKVELQTIHEEELRLQEHYSQSRLCLQDTLCLHQG PHPKALPAPAPNWQATQGEADSMGLRERIQELEAQMDVMREELGHKDLGDAATLREKYQRD LESLKATCERGFAAMEETHQKKIEDLQRQHRELEKLREEKDRLLAEETAATISAIEMKNAH REEMERELEKSQRSQISSVNSDVEALRRQYLEELQSVQRELEVLSQYSQKLENLAHAQALE AERQALRQCQRENQELNAHNQELNNRLAAEITRLRTLTTGDDGGGATGSPLAQGKDAYELEV LRVKESEIQYLKQEISSLKDELQALRDKKYASDKYKDIYELSLIAKAKADCDISRLKEQLKA ATEALGEKSPDSATVSGYDIMKSKSNPDFLKKDRSCVTRQLRNIRSKSLKEGLTVQERLKLFE SRDLKKD		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 54B.

Table 54B. Comparison of NOV54a against NOV54b.		
Protein Sequence	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV54b	710..1019 939..1248	240/310 (77%) 244/310 (78%)

Further analysis of the NOV54a protein yielded the following properties shown in Table 54C.

Table 54C. Protein Sequence Properties NOV54a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV54a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 54D.

Table 54D. Geneseq Results for NOV54a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB57200	Mouse ischaemic condition related protein sequence SEQ ID NO:488 - Mus musculus, 1024 aa. [WO200188188-A2, 22-NOV-2001]	1..1024 1..1023	937/1027 (91%) 967/1027 (93%)	0.0
AAM40793	Human polypeptide SEQ ID NO 5724 - Homo sapiens, 612 aa. [WO200153312-A1, 26-JUL-2001]	347..919 1..573	557/573 (97%) 559/573 (97%)	0.0
ABG10511	Novel human diagnostic protein #10502 - Homo sapiens, 448 aa. [WO200175067-A2, 11-OCT-2001]	591..1019 1..429	419/429 (97%) 423/429 (97%)	0.0
ABG10512	Novel human diagnostic protein #10503 - Homo sapiens, 1201 aa. [WO200175067-A2, 11-OCT-2001]	710..1019 871..1182	286/312 (91%) 291/312 (92%)	e-152
AAB41888	Human ORFX ORF1652 polypeptide sequence SEQ ID NO:3304 - Homo sapiens, 233 aa. [WO200058473-A2, 05-OCT-2000]	342..574 1..233	232/233 (99%) 232/233 (99%)	e-137

10

In a BLAST search of public sequence databases, the NOV54a protein was found to have homology to the proteins shown in the BLASTP data in Table 54E.

Table 54E. Public BLASTP Results for NOV54a				
Protein Accession Number	Protein/Organism/Length	NOV54a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
CAD39169	Hypothetical protein - Homo sapiens (Human), 987 aa (fragment).	38..1025 1..987	984/988 (99%) 984/988 (99%)	0.0
P97434	Rho-interacting protein 3 (p116RIP) (RIP3) - Mus musculus (Mouse), 1024 aa.	1..1024 1..1023	937/1027 (91%) 967/1027 (93%)	0.0
Q9ERE6	Rho-interacting protein 3 (p116RIP) (RIP3) - Rattus norvegicus (Rat), 1029 aa.	1..1024 1..1028	934/1032 (90%) 968/1032 (93%)	0.0
Q96G40	Unknown (Protein for IMAGE:4121355) - Homo sapiens (Human), 845 aa (fragment).	156..1019 3..826	820/864 (94%) 822/864 (94%)	0.0
BAC03851	CDNA FLJ34968 fis, clone NTONG2004844, highly similar to P116 RHO-INTERACTING PROTEIN - Homo sapiens (Human), 586 aa.	208..793 1..586	585/586 (99%) 585/586 (99%)	0.0

Pfam analysis predicts that the NOV54a protein contains the domains shown in the Table 54F.

Table 54F. Domain Analysis of NOV54a			
Pfam Domain	NOV54a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PH	44..145	24/102 (24%) 82/102 (80%)	1.9e-12
PH	388..483	31/97 (32%) 73/97 (75%)	3.1e-20

5

Example 55.

The NOV55 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 55A.

Table 55A. NOV55 Sequence Analysis			
	SEQ ID NO: 207	669 bp	
NOV55a,	ATGGCGGATGGGAGCAGCGATGCGGCTAGGGAACCTCGCCCTGCACCAGCCCCAATCAGACGC		

311

	TGAGCACCCGCTGCCAGTCGCTGGAGTTGGCCGGGCTGGGCTTCGCGGAGCTGCAGGACTTGT GCCGACAGCTCCACGCCCGTGTGGACAAGGTGGATGAAGAGAGATACGACATAGAGGCAAAAG TCACCAAGAACATCACGAAGATCTTTGACCTTCGAGGCAAGTTTAAGCGGCCACCCTGCGGA GAGTGAGGATCTCTGAGATGCCATGATGCAGGCGCTGCTGGGGGCCCGGGCTAAGGAGTCCC TGGACCTGCGGGCCACCTCAAGCAGGTGAAGAAGGAGGACACCGAGAAGGAAAAACCGGAGG TGGGAGACTGGCGCAAGAACATCGATGCACTGAGTGAATGGAGGGCCGCAAGAAAAAGTTTG AGAGCTGAGCCTTCTGCTACTGCCCTGCCCTGAGGAGGGCCCTGAGGAATAAAGCTTCTC TCTGAGCTGAAAAAAAAAAAA		
	ORF Start: ATG at 90		ORF Stop: TGA at 699
	SEQ ID NO: 214	203 aa	MW at 23226.3kD
NOV55d, CG132650-03 Protein Sequence	MADGSSDAAREPRPAPAPIRRRSSNYRAYATEPHAKKSKISASRKLQLKTLTLLQIAKQELER EAEERRGEKGRALSTRCQSLAGLGFELQDLRQLHARVDKVDEERYDIEAKVTKNITKIF DLRGKFKRPTLRRVRI SADAMMQALLGARAKESLDLRAHLKQVKKEDTEKENREVGDRKNID ALSGMEGRKKKFES		
	SEQ ID NO: 215	855 bp	
NOV55e, CG132650-04 DNA Sequence	CGGCCGCGTCGACCCGGAGGAGACTGACGGTCCCTGGGACCTGAAGGTACCCGGGCGGCC CCTCACTGACCTCCAAACGCCCTGTCTCGCCCTGCCTCCTGCCATTCCCGGCCTGAGTCT CAGCATGGCGGATGGGAGCAGCGATGCGGCTAGGGAACCTCGCCCTGCACCGCCCAATCAG ACGCCGCTCCTCCAACACCGCGCTTATGCCACGGAGCCGACGCCAAGAAAAAATCTAAGAT CTCCGCTCGAGAAAATTGCAGCTGAAGACTCTGCTGCTGCAGATTGCAAGCAAGAGCTGGA GCGAGAGGCGGAGGAGCGGCGGAGAGAAGGGGCGCGCTCTGAGCACCCGCTGCCAGCCGCT GGAGTTGGCCGGGCTGGGCTTCGCGGAGCTGCAGGACTTGTGCCACAGCTCCACGCCCGTGT GGACAAGGTGGATGAAGAGAGATACGACATAGAGGCAAAAGTCACCAAGAACATCACGGAGAT TGCAGATCTGACTCAGAAGATCTTTGACCTTCGAGGCAAGTTTAAGCGGCCACCCTGCGGAG AGTGAGGATCTCTGCAGATGCCATGATGCAGGCGCTGCTGGGGGCCCGGGCTAAGGAGTCCCT GGACCTGCGGGCCACCTCAAGCAGGTGAAGAAGGAGGACACCGAGAAGGAAAAACCGGAGGT GGGAGACTGGCGCAAGAACATCGATGCACTGAGTGAATGGAGGGCCGCAAGAAAAAGTTTGA GAGCTGAGCCTTCTGCTACTGCCCTGCCCTGAGGAGGGCCCTGAGGAATAAAGCTTCTCT CTGAGCTGAAAAAAAAAAAAAAAAAACCACAAAAA		
	ORF Start: ATG at 131		ORF Stop: TGA at 761
	SEQ ID NO: 216	210 aa	MW at 24007.2kD
NOV55e, CG132650-04 Protein Sequence	MADGSSDAAREPRPAPAPIRRRSSNYRAYATEPHAKKSKISASRKLQLKTLTLLQIAKQELER EAEERRGEKGRALSTRCQSLAGLGFELQDLRQLHARVDKVDEERYDIEAKVTKNITEIA DLTQKIFDLRGKFKRPTLRRVRI SADAMMQALLGARAKESLDLRAHLKQVKKEDTEKENREVG DWRKNIDALSGMEGRKKKFES		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 55B.

Table 55B. Comparison of NOV55a against NOV55b through NOV55e.		
Protein Sequence	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV55b	1..210 1..203	154/210 (73%) 154/210 (73%)
NOV55c	69..210 27..168	142/142 (100%) 142/142 (100%)
NOV55d	1..210 1..203	153/210 (72%) 153/210 (72%)
NOV55e	1..210 1..210	161/210 (76%) 161/210 (76%)

Further analysis of the NOV55a protein yielded the following properties shown in Table 55C.

Table 55C. Protein Sequence Properties NOV55a	
PSort analysis:	0.9855 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV55a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 55D.

Table 55D. Geneseq Results for NOV55a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW41570	Modified human cardiac troponin I HcTnI-K6-H5-D - Homo sapiens, 222 aa. [WO9739132-A1, 23-OCT-1997]	1..222 1..222	222/222 (100%) 222/222 (100%)	e-124
AA Y03174	Recombinant human troponin I - Homo sapiens, 226 aa. [WO9854219-A1, 03-DEC-1998]	1..221 8..225	214/221 (96%) 214/221 (96%)	e-117
AA Y03168	Recombinant Human cardiac troponin I - Homo sapiens, 226 aa. [WO9854218-A1, 03-DEC-1998]	1..221 8..225	214/221 (96%) 214/221 (96%)	e-117
AAW18054	Recombinant human myofibrillar contractile protein Troponin I - Synthetic, 226 aa. [WO9719955-A1, 05-JUN-1997]	1..221 8..225	214/221 (96%) 214/221 (96%)	e-117
AAW41573	Modified human cardiac troponin I HcTnI-(HL)3 - Homo sapiens, 216 aa. [WO9739132-A1, 23-OCT-1997]	1..221 1..215	213/221 (96%) 213/221 (96%)	e-115

- 10 In a BLAST search of public sequence databases, the NOV55a protein was found to have homology to the proteins shown in the BLASTP data in Table 55E.

Table 55E. Public BLASTP Results for NOV55a				
Protein Accession Number	Protein/Organism/Length	NOV55a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
TPHUIC	troponin I, cardiac muscle - human, 210 aa.	1..210 1..210	210/210 (100%) 210/210 (100%)	e-114
P19429	Troponin I, cardiac muscle - Homo sapiens (Human), 209 aa.	2..210 1..209	209/209 (100%) 209/209 (100%)	e-114
AAM33343	Cardiac troponin I - Canis familiaris (Dog), 211 aa.	1..209 1..210	200/210 (95%) 203/210 (96%)	e-107
I56441	troponin I - rat, 211 aa.	1..209 1..210	195/210 (92%) 202/210 (95%)	e-106
A53805	troponin I, cardiac - mouse, 211 aa.	1..209 1..210	195/210 (92%) 201/210 (94%)	e-106

PFam analysis predicts that the NOV55a protein contains the domains shown in the Table 55F.

Table 55F. Domain Analysis of NOV55a			
Pfam Domain	NOV55a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Troponin	46..177	77/190 (41%) 123/190 (65%)	8.6e-59

5

Example 56.

The NOV56 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 56A.

Table 56A. NOV56 Sequence Analysis			
	SEQ ID NO: 217	545 bp	
NOV56a, CG133808-01 DNA Sequence	GGCGACTGGCTGCACGCGCCGCTGGAGCCCGACGTGCGCGGGTGGTGGGCTTTGACCCG CACTTCAGCTACATGAAGCTCACCAAGGCCCTGCGCTACCTGCAGCAGCCCGGCTGCCTGCTC GTGGGCACCAACATGGACAACCGGCTTCCGCTTGAGAACGGCCGCTTCATCGCGGGTACCGGG TGTCTGGTCCGAGCCGTGGAGATGGCCGCCAGCGCCAGCCGACATCATCGGGAAGCCAGC CGCTTCATTTTCGACTGCGTGTCCAGGAATACGGCATCAACCCGAGCGCACCGTCATGGTG GGAGACCGCCTGGACACAGACATCCTCCTAGGCGCCACCTGTGGCCTGAAGACCATCCTGACC CTCACCGAGTCTCCACTCTAGGGGATGTGAAGAATAATCAGGAAAGTGACTGCGTGTCTAAG AAGAAATGGTCCCTGACTTCTATGTTGACAGCATAGCCGACCTTTGCGCTGCCCTTCAAGGT TAAAGATTGAGTGTCTTTAATCTGCAGAATAAAAAAAAAA		
	ORF Start: ATG at 76		ORF Stop: TAA at 505
	SEQ ID NO: 218	143 aa	MW at 15565.9kD

NOV56a, CG133808-01 Protein Sequence	MKLTKALRYLQOPGCLLVGTNMDNRLPLENGRFIAGTGCLVRVEMAAQRQADIIGKPSRFIF DCVSQEYGINPERTVMVGDRDLTDILLGATCGLKTLTLTGVS TLGDVKNNQESDCVSKKKMV PDFYVDSIADLLPALQG
--	--

Further analysis of the NOV56a protein yielded the following properties shown in Table 56B.

Table 56B. Protein Sequence Properties NOV56a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.2184 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV56a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 56C.

Table 56C. Geneseq Results for NOV56a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAE14458	Human protein phosphatase-8 - Homo sapiens, 321 aa. [WO200196546-A2, 20-DEC-2001]	1..143 179..321	143/143 (100%) 143/143 (100%)	2e-78
AAM04137	Peptide #2819 encoded by probe for measuring breast gene expression - Homo sapiens, 94 aa. [WO200157270-A2, 09-AUG-2001]	37..130 1..94	94/94 (100%) 94/94 (100%)	5e-49
AAM28902	Peptide #2939 encoded by probe for measuring placental gene expression - Homo sapiens, 94 aa. [WO200157272-A2, 09-AUG-2001]	37..130 1..94	94/94 (100%) 94/94 (100%)	5e-49
AAM16403	Peptide #2837 encoded by probe for measuring cervical gene expression - Homo sapiens, 94 aa. [WO200157278-A2, 09-AUG-2001]	37..130 1..94	94/94 (100%) 94/94 (100%)	5e-49
AAM68596	Human bone marrow expressed probe encoded protein SEQ ID NO: 28902 - Homo sapiens, 94 aa. [WO200157276-A2, 09-AUG-2001]	37..130 1..94	94/94 (100%) 94/94 (100%)	5e-49

10

In a BLAST search of public sequence databases, the NOV56a protein was found to have homology to the proteins shown in the BLASTP data in Table 56D.

Table 56D. Public BLASTP Results for NOV56a				
Protein Accession Number	Protein/Organism/Length	NOV56a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CVY8	1700012G19Rik protein - Mus musculus (Mouse), 224 aa (fragment).	1..143 82..224	134/143 (93%) 139/143 (96%)	1e-72
Q9D6Q5	1700012G19Rik protein - Mus musculus (Mouse), 122 aa.	22..143 1..122	115/122 (94%) 119/122 (97%)	5e-61
Q8VD52	Reg I binding protein I - Rattus norvegicus (Rat), 204 aa (fragment).	2..143 49..187	68/142 (47%) 96/142 (66%)	3e-31
Q9UGY2	DJ37E16.5 (Novel protein similar to NITROPHENYLPHOSPHATASES from VARIOUS ORGANISMS) (Hypothetical 31.7 kDa protein) - Homo sapiens (Human), 296 aa.	2..142 158..295	68/141 (48%) 94/141 (66%)	2e-30
Q96GD0	Similar to hypothetical protein dJ37E16.5 - Homo sapiens (Human), 179 aa (fragment).	2..142 41..178	68/141 (48%) 94/141 (66%)	2e-30

- 5 PFam analysis predicts that the NOV56a protein contains the domains shown in the Table 56E.

Table 56E. Domain Analysis of NOV56a			
Pfam Domain	NOV56a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 57.

- 10 The NOV57 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 57A.

Table 57A. NOV57 Sequence Analysis			
	SEQ ID NO: 219	339 bp	
NOV57a, CG136288-01 DNA Sequence	CCTGCCCACTTTACCCAGATGTCCAGCAAGGTGGCCATCAACAGTGACATTGGGCAGGCCCTC TGGGCAGTGGAGCAGCTCCAGATGGAGGCAGGCATCGACCAAGTGAAGGTGAGGAAGATGGCC GCCGATCTGCTGAAGTTCTGCACGGAGCAGGCCAAGAATGACCCCTTCCTTGTGGGCATCCCG GCCGCCACCAACTCCTTCAAGGAGAAGAAGCCCTATGCCATCCTATGAACCCAGGGCAATGCC		

	ACCCTGTGGCCTGGGCAAACCAGGGGCCTCAATAAACATGAAGTGAATACTTCTCAGGGCAT GGCTGAGCTGGGCTGAGATGGGAG		
	ORF Start: ATG at 19		ORF Stop: TGA at 235
	SEQ ID NO: 220	72 aa	MW at 7911.2kD
NOV57a, CG136288-01 Protein Sequence	MSSKVAINSDIGQALWAVEQLQMEAGIDQVKVRKMAADLLKFCTEQAKNDPFLVGIPAAATNSF KEKKPYAIL		

Further analysis of the NOV57a protein yielded the following properties shown in Table 57B.

Table 57B. Protein Sequence Properties NOV57a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV57a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 57C.

Table 57C. Geneseq Results for NOV57a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB33831	Human secreted protein BLAST search protein SEQ ID NO: 175 - Homo sapiens, 72 aa. [WO200056753-A1, 28-SEP-2000]	1..72 1..72	38/72 (52%) 52/72 (71%)	1e-15
AAO14753	Human Guanine nucleotide binding protein gamma 7 (GNG7) - Homo sapiens, 68 aa. [WO200218647-A1, 07-MAR-2002]	5..72 1..68	35/68 (51%) 52/68 (76%)	4e-15
AAW09416	Human G protein gamma-7 subunit - Homo sapiens, 68 aa. [WO9637513-A1, 28-NOV-1996]	5..72 1..68	35/68 (51%) 52/68 (76%)	4e-15
ABB57442	Human secreted protein encoding polypeptide SEQ ID NO 88 - Homo sapiens, 72 aa. [WO200183510-A1, 08-NOV-2001]	1..72 1..72	36/72 (50%) 52/72 (72%)	7e-15

ABB57441	Human secreted protein encoding polypeptide SEQ ID NO 87 - Homo sapiens, 72 aa. [WO200183510-A1, 08-NOV-2001]	1..72 1..72	36/72 (50%) 52/72 (72%)	7e-15
----------	---	----------------	----------------------------	-------

In a BLAST search of public sequence databases, the NOV57a protein was found to have homology to the proteins shown in the BLASTP data in Table 57D.

Table 57D. Public BLASTP Results for NOV57a				
Protein Accession Number	Protein/Organism/Length	NOV57a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q28024	Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-12 subunit (Gamma-S1) - Bos taurus (Bovine), 71 aa.	2..72 1..71	37/71 (52%) 51/71 (71%)	1e-14
O60262	Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-7 subunit - Homo sapiens (Human), 68 aa.	5..72 1..68	35/68 (51%) 52/68 (76%)	1e-14
AAM12593	Guanine nucleotide binding protein gamma 12 - Homo sapiens (Human), 72 aa.	1..72 1..72	36/72 (50%) 52/72 (72%)	2e-14
AAK61365	G-protein gamma 7 - Mus musculus (Mouse), 68 aa.	9..72 5..68	33/64 (51%) 50/64 (77%)	8e-14
P30671	Guanine nucleotide-binding protein G(I)/G(S)/G(O) gamma-7 subunit (Gamma-II) - Bos taurus (Bovine), 68 aa.	5..72 1..68	33/68 (48%) 52/68 (75%)	8e-14

5

PFam analysis predicts that the NOV57a protein contains the domains shown in the Table 57E.

Table 57E. Domain Analysis of NOV57a			
Pfam Domain	NOV57a Match Region	Identities/ Similarities for the Matched Region	Expect Value
G-gamma	11..65	23/55 (42%) 47/55 (85%)	5.6e-17

Example 58.

The NOV58 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 58A.

Table 58A. NOV58 Sequence Analysis			
	SEQ ID NO: 221	1332 bp	
NOV58a, CG136933-01 DNA Sequence	ATAATGGATTCTGAATTAATGCATAGTATAGTAGGAAGCTATCATAAACCTCCAGAAAGAGTA TTTGTTCCCTCATTACCCAGAATGAACCATCTCAGAATTGCCATCCTGCAAACCTAGAAAGTT ACCTCTCCTAAGATACTTCTATAGCCCAATAGCCAAGCTCTTATTTAGCCTTAAAACTCTT CAGGAAAAAATTCATCGTTTAGAGCTGGAGAGAACAAGCTGAAGATAACCTGAACATTCTT TCCAGAGAAGCAGCACAGTATAAGAAGGCCTTAGAGAATGAAACAAATGAGAGAAATCTGGCA CACCAGGAGCTGATAAAGCAGAAAAAGATATAAGTATACAGTTAAGCTCAGCCAGTCTCGT TGTACTCTTCTAGAGAAGCAACTAGAATATACAAAGAGAATGGTTCTCAACGTAGAGCGAGAA AAGAACATGATCCTAGAACAACAGGCCAGCTTCAGAGGGAAAAAGAACAGATCAGATGAAG CTGTATGCAAAACTTGAAAAGCTTGATGTCTTAGAAAAAGAGTGTTCAGACTTACAACAAT CAGAAAACTGCTGAGGACAAGATTAAACATTTAGAAGAAAACTTAAGGAAGAAGAACATCAG CGTAAGCTATTTCAAGACAAAGCTTCTCAGCTTCAAACCTGGACTTGAATCAGTAAATATATA ATGCTTTCAGTTTCAAATCTAAAGCACTCCAAGGAAAAAGAAATCTTCAAAGTTTTTGACG ATGAGGCAACATCGTGACCCACATATCCTTCAGAAACCTTTTAACGTGACTGAGACTAGATGT CTCCCAAGCCTTCTAGAACAATCTCCTGGTGTAAGCTATTCTCCTGACTCAGAAAAGTCC ATTTCCATTGTGACAATTTATCTGAACCTTTGATGGCAATGCAAGATGAGCTGGACCAATG AGCATGGAGCACCAAGAACTACTGAAACAAATGAAGGAACTGAAAGTCATTAGTCTGTGAC GACATAGAATGTGAAGTAGAGTGTCTTACTCAAGAAAAATGGAATTAAGGAGAACAAATCTCC AAAGTGAAGAAGCATCAAGACAGTGTAAAGAGGCTTCAGCAAAAAGTTCAAACTCAAAGATG AGTGAAGCTTCAGGTATTCAGCAAGAAGACAGCTACCCTAAAGGATCAAAGAACATAAAAAAT AGCCCCAGAAAAATGTTTACTGACACTAACCTTTTTCAGAAAAACAGCAGCTTTCATCCAATA CGAGTTTATAATCTTCAAATGAAATTGAGAAGAGATGATATCATGTGGGAACAGTAACAAAAAC AGCAAAACT		
	ORF Start: ATG at 4		ORF Stop: TAA at 1315
	SEQ ID NO: 222	437 aa	MW at 51102.1kD
NOV58a, CG136933-01 Protein Sequence	MDSELMHSIVGSYHKPPERVFVPSFTQNEPSQNCHPANLEVTSPKILHSPNSQALILALKTLQ EKIHRLELERTQAEDNLNLSREAAQYKKALENETERNLAHQELIKQKKDISIQLSSAQSRC TLLEKQLEYTKRMVLNVEREKNMILEQQAQLQREKEQDQMKLYAKLEKLDVLEKECFRLTTTQ KTAEDKIKHLEELKEEEHQKRLFQDKASQLQTGLEISKIIMSSVSNLKHSEKKKSSKFLQM RQHRDPHILQKPFNVTTETRLPKPSRTTSWCKAIPDSEKISICDNLSELLMAMQDELQDMS MEHQELLQMKETESHVSCDDIECELECLLKKMEIKGEQISKLLKHQDSVRRLLQKQVQNSKMS EASGIQQEDSYPKGSKNIKNSPRKCLDTNLFQKNSSFHPIRVHNLQMKLRRDDIMWEQ		

5

Further analysis of the NOV58a protein yielded the following properties shown in Table 58B.

Table 58B. Protein Sequence Properties NOV58a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV58a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 58C.

Table 58C. Geneseq Results for NOV58a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV58a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM00812	Human bone marrow protein, SEQ ID NO: 175 - Homo sapiens, 294 aa. [WO200153453-A2, 26-JUL-2001]	1..281 1..292	255/292 (87%) 260/292 (88%)	e-137
AAM00925	Human bone marrow protein, SEQ ID NO: 401 - Homo sapiens, 206 aa. [WO200153453-A2, 26-JUL-2001]	249..436 1..188	186/188 (98%) 187/188 (98%)	e-106
AAB84841	Human FGF Associated Protein, FAP - Homo sapiens, 500 aa. [JP2001061477-A, 13-MAR-2001]	51..402 65..462	138/398 (34%) 212/398 (52%)	1e-58
AAB28209	Novel human protein #7 - Homo sapiens, 114 aa. [WO200052165-A2, 08-SEP-2000]	110..220 3..113	51/111 (45%) 80/111 (71%)	8e-23
ABG15517	Novel human diagnostic protein #15508 - Homo sapiens, 662 aa. [WO200175067-A2, 11-OCT-2001]	54..229 451..653	68/203 (33%) 108/203 (52%)	2e-20

5

In a BLAST search of public sequence databases, the NOV58a protein was found to have homology to the proteins shown in the BLASTP data in Table 58D.

Table 58D. Public BLASTP Results for NOV58a				
Protein Accession Number	Protein/Organism/Length	NOV58a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9CWM5	2410017P07Rik protein - Mus musculus (Mouse), 429 aa.	1..437 1..429	309/437 (70%) 357/437 (80%)	e-169
Q8VDS7	Similar to RIKEN cDNA 2410017P07 gene - Mus musculus (Mouse), 400 aa.	1..437 1..400	289/437 (66%) 339/437 (77%)	e-153
Q9CZE0	2410017P07Rik protein - Mus musculus (Mouse), 353 aa.	130..437 50..353	206/308 (66%) 246/308 (78%)	e-110
Q9DSA1	4930484D11Rik protein - Mus musculus (Mouse), 256 aa.	1..248 1..244	203/248 (81%) 219/248 (87%)	e-108

Q14704	KIAA0092 protein - Homo sapiens (Human), 474 aa.	51..402 65..436	137/375 (36%) 212/375 (56%)	8e-60
--------	--	--------------------	--------------------------------	-------

PFam analysis predicts that the NOV58a protein contains the domains shown in the Table 58E.

Table 58E. Domain Analysis of NOV58a			
Pfam Domain	NOV58a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 59.

The NOV59 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 59A.

Table 59A. NOV59 Sequence Analysis			
	SEQ ID NO: 223	1045 bp	
NOV59a, CG136942-01 DNA Sequence	<p>TCTCCATGACATGTTGATGCTGGCTGAGCAGCAGCAGAAGCAGAAGTGGGCTGTGAATACTCA AAACACTGCCTGGAGTAATGCTGATTCTAAATTTGGCCAGAGGATACTAGAGAAGATGGAATG GTCTAAAGGAAGGGGTTTAGGGGTTTCAGGAGCAAGGAGGCCAGATGATATTAAAGTTCAAGT TAAAAATAACGACCTGGGACTTCAAGCTACAATCAATAATGAAGCCAAGTGGATTGCCCATCA AGATGATTTTAACTGGCTTCTGGCGGAAGTGAACACTTGTTCAGAGGCAGGAAACAGCAGACTC CTTAGACAACAAGAAAAAGAAATATTTTAGTCTTGAAGAAATTTCCAAAATCTTCAAAAAGT TGTTTCATCATAGGAAATTTACAAAAGAAAAGGATCTATCATCTCGGAGCAAAACAGATCGTGA CTGCATTTTGGGAAAAACAGAGTAAGAAGACTCCCGAGGGTAATTCAGTCCCTCCACTCC AGACCAGAACAAAACACGATGACAACCCATGCCTTCACCATCCAGGAGCGTTTGGCAAGCG AATGGCAGCACTGAAGAACAAGCCCCAGGTTGCAGCTCCAGGGCTGACATTTCAGAGCCCA AGTGGAAATGCAAAAGGGGGAAGAAAAGAAACAAAGAGGCAACAGGTAAAAATGGGGAGAGTTA CCCCCAACACAGCCTAAGGCCAAGCGGCCTAAAGAGGGAAAGCCTAAGAGAGACAAGGTCCA GAAGTCGGCATCCAAGGAGAAAAGAGCACGGACAGACGGACAGTGCAGAGGCCTCTGCTGGGA AGAGAGTTCTGAGGCCTCTGCTCAGGTTGCAGGGAATGTGTGCAGCCACCTGATGGCCAGGA TTTCACCCTGAAGCCCAAAAAGACAAGAGGAAAAAAGCTGCAAGCCAGTAGAGGTAGC AATGGACACTACGCTGAAAGAAACACCAATGAAAAATAAGAAAAAGAAGAAAGGTTCCAAATG AATTCTCTCCAGCCAGGGCCTTCGACCACTCAGCTT</p>		
	ORF Start: ATG at 11		ORF Stop: TGA at 1007
	SEQ ID NO: 224	332 aa	MW at 37378.2kD
NOV59a, CG136942-01 Protein Sequence	<p>MLMLAEQQQKQKWAVNTQNTAWSNADSKFGQRILEKMEWSKGRGLGVQEQGPDDIKVQVKNN DLGLQATINNEANWIAHQDDFNWLLAELNTCQRQETADSLDNKKKKYFSLEEISKIFKNCVHH RKFTKEKDLSSRSKTDRCIFGKKQSKKTPEGNSSPSTPDQNKTTMTTHAFTIQERFAKRMMA LKNKPQVAAPGPDISKQVECKRGKKRNKEATGKNGESYPPTQPKAKRPKEGPKPRDKVQKSA SKEKRARTDQCRGLCWEESEASAQAGNVCVQPPDQDFTLKPKKTRGKKKAAPVEVAMDT TLKETPMKNKKKKKGSK</p>		

10

Further analysis of the NOV59a protein yielded the following properties shown in Table 59B.

Table 59B. Protein Sequence Properties NOV59a	
PSort analysis:	0.9748 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV59a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 59C.

5

Table 59C. Geneseq Results for NOV59a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV59a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM78624	Human protein SEQ ID NO 1286 - Homo sapiens, 328 aa. [WO200157190-A2, 09-AUG-2001]	1..332 1..328	244/332 (73%) 275/332 (82%)	e-133
AA Y32206	Human receptor molecule (REC) encoded by Incyte clone 2825826 - Homo sapiens, 352 aa. [WO9957270-A2, 11-NOV-1999]	1..329 1..326	236/329 (71%) 272/329 (81%)	e-131
AAM79608	Human protein SEQ ID NO 3254 - Homo sapiens, 322 aa. [WO200157190-A2, 09-AUG-2001]	1..286 39..321	211/286 (73%) 237/286 (82%)	e-115
ABB12307	Human NADH-cytochrome b5 reductase homologue, SEQ ID NO:2677 - Homo sapiens, 322 aa. [WO200157188-A2, 09-AUG-2001]	1..286 39..321	211/286 (73%) 237/286 (82%)	e-115
ABB12190	Human tumour suppressor homologue, SEQ ID NO:2560 - Homo sapiens, 127 aa. [WO200157188-A2, 09-AUG-2001]	1..102 5..106	102/102 (100%) 102/102 (100%)	1e-55

In a BLAST search of public sequence databases, the NOV59a protein was found to have homology to the proteins shown in the BLASTP data in Table 59D.

Table 59D. Public BLASTP Results for NOV59a				
Protein Accession Number	Protein/Organism/Length	NOV59a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96BK5	Pin2-interacting protein X1 (TRF1-interacting protein 1) (Liver- related putative tumor suppressor) (67-11-3 protein) - Homo sapiens (Human), 328 aa.	1..332 1..328	244/332 (73%) 275/332 (82%)	e-133
Q9CZX5	Pin2-interacting protein X1 (TRF1-interacting protein 1) (Liver- related putative tumor suppressor) (LPTS1) (67-11-3 protein) - Mus musculus (Mouse), 332 aa.	1..328 1..329	199/333 (59%) 238/333 (70%)	e-101
Q22705	T23G7.3 protein - Caenorhabditis elegans, 339 aa.	1..247 1..262	82/268 (30%) 131/268 (48%)	1e-27
Q9V952	CG11180 protein - Drosophila melanogaster (Fruit fly), 726 aa.	1..328 1..382	99/386 (25%) 165/386 (42%)	2e-22
Q9URX9	Hypothetical 31.9 kDa protein C890.05 in chromosome I - Schizosaccharomyces pombe (Fission yeast), 284 aa.	4..256 3..244	65/258 (25%) 116/258 (44%)	3e-11

PFam analysis predicts that the NOV59a protein contains the domains shown in the Table 59E.

Table 59E. Domain Analysis of NOV59a			
Pfam Domain	NOV59a Match Region	Identities/ Similarities for the Matched Region	Expect Value
G-patch	26..70	18/47 (38%) 33/47 (70%)	6.8e-07

5

Example 60.

The NOV60 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 60A.

Table 60A. NOV60 Sequence Analysis			
	SEQ ID NO: 225	935 bp	
NOV60a, CG137017-01 DNA	ATGTCCAACCGAGTGGTCCGCTGGGAAGCCAGCCACTCTGGGAAGTGGTACACAGCCTCAGGACACAGCTGAATGCACAGCTAGAAGGTTGGCTTTCACAAGTAAAGTCTACAAAAAGACCTGCTAGAGCCATTATTGCACCCATGCAGGATATACGTAAGTGGGTCTGTGCCACCCATGCTTAT		

Sequence	AAACAAGTGGATCCGTCTATTACCCAGGGAATTTTCATCCTTGGGCCTTCTCATCATGTGTCC CTCTCTCAGTGTGCACTTTCCAGTGTGGATATATATAGGACACCTCTGCATGACCTTCGTATT GACCAAAAGATTTACAGAGAACTGTGGAAGACAGGAATGTTTGACGCATGTCTTGCAGACA GACAAAGATGAACACGGTATTGAAATGCATTTCGCTTGTACAGCTAAAGCCATGGAAATCCAT AAGGATGAGTTTACCATTATTCCTGTACTGGTGGAGCTCTGAGTGAGTCAAAGAACAGGAA TTCGGAAGAACTCTTCAGTAAATATCTAGCAGATCCTAGTAATTCTTTGTGGTTCTTCTGAT TTTTGCCATTGGGGTCAAAGATTCCGTTACAGTTACTATGATGAATCCCAGGGGGAGACTTAT AGATCCATTGAACATCTAGATAAAATGGGTATGAGTATTATAGAACAATTAGACCCTGTATCT TTTAGCAATTACTTGAGGAACACCATAATACTATATGTGGAAGACATCCCTTTAGGGTGTTA AATGCTATCACAGAGCTCCAGAAGAAATGGAAGAAATATGAGCTTTTCCTTTTGAATTATGCC CAGTCAAGCCAGTGTAGAACTGGCAAGACAGTTCAGTGAGTTACACAAGTGGAGCGCTCAGC GTCCGCTGAAGCTCTGAATCCTCAGGAGGCCACCTGCACATTCTCATACTCT		
	ORF Start: ATG at 1		ORF Stop: TGA at 889
	SEQ ID NO: 226	296 aa	MW at 33879.9kD
NOV60a, CG137017-01 Protein Sequence	MSNRVVRWEASHSGNWTASGPQLNAQLEGWLSQVKSTKRPARAI IAPHAGYTYCGSCATHAY KQVDPSITQGIFILGPSHHVLSQCALSSVDIYRTPHDLRIDQKIYRELWKTGMFERMSLQT DKDEHGIEMHLPCTAKAMEIHKDEFTIIIPVLVGALSESKEQFGKLF SKYLADPSNFFVSSD FCHWGQRF RYSYYDESQGETYRSIEHLDKMGMSIIIEQLDPVSFSNYLRKHNHTICGRHPFRVL NAITELQKNGRNMSFSFLNYAQSSQCRNWQDSSVSYYTGALTVR		

Further analysis of the NOV60a protein yielded the following properties shown in Table 60B.

Table 60B. Protein Sequence Properties NOV60a	
PSort analysis:	0.5297 probability located in microbody (peroxisome); 0.4657 probability located in mitochondrial matrix space; 0.1652 probability located in mitochondrial inner membrane; 0.1652 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV60a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 60C.

Table 60C. Geneseq Results for NOV60a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV60a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG12611	Novel human diagnostic protein #12602 - Homo sapiens, 297 aa. [WO200175067-A2, 11-OCT-2001]	1..295 1..296	269/296 (90%) 278/296 (93%)	e-157
ABG12609	Novel human diagnostic protein #12600 - Homo sapiens, 297 aa. [WO200175067-A2, 11-OCT-2001]	1..295 1..296	269/296 (90%) 278/296 (93%)	e-157
ABG16429				e-113

	- Homo sapiens, 256 aa. [WO200175067-A2, 11-OCT-2001]	38..256	204/219 (92%)	
ABB71637	Drosophila melanogaster polypeptide SEQ ID NO 41703 - Drosophila melanogaster, 295 aa. [WO200171042- A2, 27-SEP-2001]	10..293 6..291	173/286 (60%) 215/286 (74%)	1e-99
AAG37603	Arabidopsis thaliana protein fragment SEQ ID NO: 46262 - Arabidopsis thaliana, 291 aa. [EP1033405-A2, 06- SEP-2000]	9..292 6..286	131/285 (45%) 185/285 (63%)	8e-74

In a BLAST search of public sequence databases, the NOV60a protein was found to have homology to the proteins shown in the BLASTP data in Table 60D.

Table 60D. Public BLASTP Results for NOV60a				
Protein Accession Number	Protein/Organism/Length	NOV60a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y316	Hypothetical protein CGI-27 (C21orf19-like protein) - Homo sapiens (Human), 297 aa.	1..295 1..296	270/296 (91%) 279/296 (94%)	e-158
Q91VH6	CGI-27 protein (C21orf19-like protein) - Mus musculus (Mouse), 297 aa.	1..295 1..296	268/296 (90%) 279/296 (93%)	e-157
Q9VG04	CG8031 protein (LP04475P) - Drosophila melanogaster (Fruit fly), 295 aa.	10..293 6..291	173/286 (60%) 215/286 (74%)	4e-99
Q9SIR5	At2g25280 protein (Hypothetical 32.6 kDa protein) - Arabidopsis thaliana (Mouse-ear cress), 291 aa.	9..292 6..286	131/285 (45%) 185/285 (63%)	2e-73
T34398	hypothetical protein C37C3.8 - Caenorhabditis elegans, 302 aa.	10..295 13..299	138/289 (47%) 187/289 (63%)	4e-72

5

PFam analysis predicts that the NOV60a protein contains the domains shown in the Table 60E.

Table 60E. Domain Analysis of NOV60a			
Pfam Domain	NOV60a Match Region	Identities/ Similarities for the Matched Region	Expect Value
UPF0103	9..292	95/334 (28%) 204/334 (61%)	1.7e-70

Example 61.

The NOV61 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 61A.

5

Table 61A. NOV61 Sequence Analysis			
	SEQ ID NO: 227	607 bp	
NOV61a, CG137146-01 DNA Sequence	CATTAATAATGTCTTTCATCTGTGAGTGGATCTACAATGGCTTCTACAGTGTGCTCCAGTTTC TAGGACTCTACAAGAAATCTAGAAAACCTGTCTCTTGGGTTGGGACAATGTGGATGAAACCA TTCTTCCTCATATGCTCAAAGATGGTGGATTGGGCCAACAGGCTCCAACACGACCTCTGCCAT CAGCAGAGCTGACAACTGCTAGCCTGGCTTTTACGACTTTTGATCTTCTGGGGCATAAGCAAA CACGTTGGTTTTTGAACAGTGATCTCCCAGCAACAAATGGGATTGCCTTCTGGGTTACTGTG CAGATTGTCCTCAGATCCTGGAATCCAAAGAGGAGCTCAATGCTTTAATGACTGATGAAAGAA TATCCCATGTGCCAGTCCCTATCTTGATTAGCAAATGGACAGAACAGGCACAATCAGTGAAG AAAAATCTGTCTGCCATTTGGTCTTTATGGACAGACCACAGGAAAGGGAACCTGTGACCCTG AAGGAGCCAAATGCCTTTTCCAAGGAATCGTTCACGTTCACTGTGTACAACAACAGGGCTATG GCAAGGGCTTCTGCTGGTTTGGCCAGTATATTGACTGATG		
	ORF Start: ATG at 9		ORF Stop: TGA at 603
	SEQ ID NO: 228	198 aa	MW at 22153.3kD
NOV61a, CG137146-01 Protein Sequence	MSFICEWIYNGFYSVLQFLGLYKRSRKLVLGWDNVDETILPHMLKDGGLGQQAPTRPLPSAE LTTASLAFTTFDLLGHKQTRFWNSDLPATNGIAFLGYCADCPQILESKEELNALMTDERISH VPVLILISKLDRTGTISEEKLCLPFGLYGQTTGKGTCDPEGAKCLFQGIVHVQCVCQQGYGKG FCWFAQYID		

Further analysis of the NOV61a protein yielded the following properties shown in Table 61B.

Table 61B. Protein Sequence Properties NOV61a	
PSort analysis:	0.7480 probability located in microbody (peroxisome); 0.1830 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space; 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV61a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 61C.

Table 61C. Geneseq Results for NOV61a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV61a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG74282	Human colon cancer antigen protein SEQ ID NO:5046 - Homo sapiens, 201 aa. [WO200122920-A2, 05-APR-2001]	1..198 4..201	120/198 (60%) 138/198 (69%)	4e-59
AAU27771	Human full-length polypeptide sequence #96 - Homo sapiens, 198 aa. [WO200164834-A2, 07-SEP-2001]	1..198 1..198	120/198 (60%) 138/198 (69%)	4e-59
AAB56779	Human prostate cancer antigen protein sequence SEQ ID NO:1357 - Homo sapiens, 201 aa. [WO200055174-A1, 21-SEP-2000]	1..198 4..201	120/198 (60%) 138/198 (69%)	4e-59
ABB57346	Mouse ischaemic condition related protein sequence SEQ ID NO:970 - Mus musculus, 198 aa. [WO200188188-A2, 22-NOV-2001]	1..198 1..198	118/198 (59%) 136/198 (68%)	3e-57
AAB74777	Human SAR1 protein SEQ ID NO:4 - Homo sapiens, 198 aa. [CN1274727-A, 29-NOV-2000]	1..198 1..198	115/198 (58%) 138/198 (69%)	1e-56

In a BLAST search of public sequence databases, the NOV61a protein was found to have homology to the proteins shown in the BLASTP data in Table 61D.

5

Table 61D. Public BLASTP Results for NOV61a				
Protein Accession Number	Protein/Organism/Length	NOV61a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAM69363	GTP-binding protein Sara - Homo sapiens (Human), 198 aa.	1..198 1..198	120/198 (60%) 139/198 (69%)	4e-59
Q99JZ4	SAR1 protein - Mus musculus (Mouse), 198 aa.	1..198 1..198	120/198 (60%) 138/198 (69%)	9e-59
Q9NR31	GTP-binding protein SAR1a (COPII-associated small GTPase) - Homo sapiens (Human), 198 aa.	1..198 1..198	120/198 (60%) 138/198 (69%)	1e-58
Q9QVY2	SAR1B protein promoting vesicle budding from the endoplasmic reticulum - Cricetus griseus (Chinese hamster), 198 aa.	1..198 1..198	118/198 (59%) 137/198 (68%)	4e-58

P36536	GTP-binding protein SAR1a - Mus musculus (Mouse), 198 aa.	1..198 1..198	118/198 (59%) 136/198 (68%)	8e-57
--------	---	------------------	--------------------------------	-------

PFam analysis predicts that the NOV61a protein contains the domains shown in the Table 61E.

Table 61E. Domain Analysis of NOV61a			
Pfam Domain	NOV61a Match Region	Identities/ Similarities for the Matched Region	Expect Value
arf	8..198	55/193 (28%) 130/193 (67%)	1e-12

5

Example 62.

The NOV62 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 62A.

Table 62A. NOV62 Sequence Analysis			
	SEQ ID NO: 229	545 bp	
NOV62a, CG137566-01 DNA Sequence	GGAACGTGGCTGGTTGGAGGAGGTAGATCACCTTTCTGCGGGGACGATTTCTGTCGGTGGTA GGCTGCTACCATGAGGTTGAATCAGAACACCTTGCTGCTGGGGAAGAAGGTGGTCCTTGTAAC CTACACCTCGGAGCATGTGCCAGGTACCACGAGTGGATGAAATCAGAGGAGCTGCAGCGTTT GACAGCCTCGGAGCCGCTGACCTTGGAGCAGGAGTATGCCATGCAGTGCAGCTGGCAGGAAGA TGCAGACAAGTGACCTTATTGTGCTGGATGCCGAGAAGTGGCAGGCCAGCCAGGCGCCAC CGAAGAGAGCTGCATGGTGGGAGATGTGAACCTCTTCTCACAGATCTAGAAGACCTCACCTT GGGGGAGATCGAGGTGATGATTGCAGAGCCAGCTGCAGGGGTAAGGGCCTTGGCACTGAGGC CGTTCTCGCATGCTGTCTTACGAACTTCACTTTGAGCAGGTGGCTACGAGCAGTGTTTTC AGGAGGTGACCCTCAGACTGACAGTGAGTGAGTCCGAGCAT		
	ORF Start: ATG at 74		ORF Stop: TGA at 476
	SEQ ID NO: 230	134 aa	MW at 15130.1kD
NOV62a, CG137566-01 Protein Sequence	MRLNQNTLLGKKVVLVPYTSHEVPRYHEWMKSEELQRLTASEPLTLEQYAMQCSWQEDADK CTFIVLDAEKWQAQPGATEESCMVGDVNLFLTDLEDLTLEIEVMIAEPSCRGKGLGTEAVLA MLSYETSL		
	SEQ ID NO: 231	709 bp	
NOV62b, CG137566-02 DNA Sequence	AGGCTGCTACCATGAGGTTGAATCAGAACACCTTGCTGCTGGGGAAGAAGGTGGTCCTTGTA CCTACACCTCGGAGCATGTCCCCAGCAGGTACCACGAGTGGATGAAATCAGAGGAGCTGCAGC GTTGGACAGCCTCGGAGCCGCTGACCTTGGAGCAGGAGTATGCCATGCAGTGCAGCTGGCAGG AAGATGCAGACAAGTGACCTTCAATTGTGCTGGATGCCGAGAAGTGGCAGGCCAGCCAGGCG CCACCGAAGAGAGCTGCATGGTGGGAGATGTGAACCTCTTCTCACAGATCTAGAAGACCTCA CCTTGGGGGAGATCGAGGTGATGATTGCAGAGCCAGCTGCAGGGGTAAGGGCCTTGGCACTG AGGCCGTTCTCGCATGCTGTCTTACGGAGTGACCAGCTAGGTCTGACCAAGTTTGAGGCTA AAATTGGGCAAGGAAATGAACCAAGCATCCGGATGTTCCAGAAACTTCACTTTGAGCAGGTGG CTACGAGCAGTGTTTTTCAGGAGGTGACCTCAGACTGACAGTGAGTGAGTCCGAGCATCAGT GGCTTCTGGAGCAGACCAGCCAGCTGGAAGAGAAGCCTTACAGAGATGGGTGGGCGCAGCCCT GCTGATGGCTGGGCCTTGTGGGAGCCACTCTGTGTGAGCAGGGTGTGGGCCCATACACTTC AAAGACCAGAGCCCTG		
	ORF Start: ATG at 12		ORF Stop: TGA at 633

	SEQ ID NO: 232	207 aa	MW at 23361.3kD
NOV62b, CG137566-02 Protein Sequence	MRLNQNTLLLGKKVVLVPYTSEHVPSRYHEWMKSEELQRLTASEPLTLEQEYAMQCSWQEDAD KCTFIVLDAEKWQAQPGATEESCMVGDVNLFLTDLEDLTIGEIEVMIAEPSCRKGGLGTEAVL AMLSYGVTTLGLTKFEAKIGQGNPSIRMFQKLHFEQVATSSVFQEVTLRLTVSESEHQWLL QTSHVEEKPYRDGSAEPC		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 62B.

Table 62B. Comparison of NOV62a against NOV62b.		
Protein Sequence	NOV62a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV62b	1..130	130/131 (99%)
	1..131	130/131 (99%)

5

Further analysis of the NOV62a protein yielded the following properties shown in Table 62C.

Table 62C. Protein Sequence Properties NOV62a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV62a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 62D.

Table 62D. Geneseq Results for NOV62a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV62a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM39602	Human polypeptide SEQ ID NO 2747 - Homo sapiens, 206 aa. [WO200153312-A1, 26-JUL-2001]	1..130 1..130	130/130 (100%) 130/130 (100%)	6e-72
AAU23355	Novel human enzyme polypeptide #441 - Homo sapiens, 209 aa. [WO200155301-A2, 02-AUG-2001]	1..130 4..133	130/130 (100%) 130/130 (100%)	6e-72
AAM41388				1e-70

	- Homo sapiens, 210 aa. [WO200153312-A1, 26-JUL-2001]	4..134	130/131 (99%)	
AAB41504	Human ORFX ORF1268 polypeptide sequence SEQ ID NO:2536 - Homo sapiens, 207 aa. [WO200058473-A2, 05-OCT-2000]	1..130 1..131	129/131 (98%) 129/131 (98%)	1e-69
AAG03244	Human secreted protein, SEQ ID NO: 7325 - Homo sapiens, 89 aa. [EP1033401-A2, 06-SEP-2000]	1..88 1..88	88/88 (100%) 88/88 (100%)	2e-47

In a BLAST search of public sequence databases, the NOV62a protein was found to have homology to the proteins shown in the BLASTP data in Table 62E.

Table 62E. Public BLASTP Results for NOV62a				
Protein Accession Number	Protein/Organism/Length	NOV62a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9BTD0	DKFZP564C103 protein - Homo sapiens (Human), 206 aa.	1..130 1..130	130/130 (100%) 130/130 (100%)	2e-71
Q9Y3T3	Hypothetical 23.3 kDa protein - Homo sapiens (Human), 206 aa.	1..130 1..130	129/130 (99%) 129/130 (99%)	1e-70
Q9BTE0	DKFZP564C103 protein - Homo sapiens (Human), 207 aa.	1..130 1..131	130/131 (99%) 130/131 (99%)	4e-70
Q9D151	I110028N05Rik protein (RIKEN cDNA I110028N05 gene) - Mus musculus (Mouse), 241 aa.	1..130 1..130	109/130 (83%) 117/130 (89%)	2e-58
Q9D7G2	I110028N05Rik protein - Mus musculus (Mouse), 211 aa.	31..130 1..100	82/100 (82%) 87/100 (87%)	3e-40

5

PFam analysis predicts that the NOV62a protein contains the domains shown in the Table 62F.

Table 62F. Domain Analysis of NOV62a			
Pfam Domain	NOV62a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 63.

The NOV63 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 63A.

Table 63A. NOV63 Sequence Analysis			
	SEQ ID NO: 233	683 bp	
NOV63a, CG137707-01 DNA Sequence	GGTGTGCTGCCAGAGATTTTGCCCTCTTCAAGGTGAATGCGGCTTCAAGGGGCTATCTTTGTGC TCCTGCCCCACCTGGGGCCCATCCTGGTCTGGCTGTTCACTCGTGATCACATGTCTGGTTGGT GTGAGGGCCCGAGGATGCTGTCCTGGTGCCCATTTCTACAAAGTCTTATTGCTTGTACAGACAG CCATCTACTCTGTCGTGGGGTATGCCTCCTACCTGGTGTGGAAGGACCTGGGAGGGGGCTTGG GGTGGCCCTGGCCCTGCCTCTTGCCCTCTATGCTGTTTACAGCTCACCATCAGCTGGACTGTCC TGGTTCTCTTTTTCACAGTCCACAACCCCGGCTCTATGCCAGGCCCTGCTGCACCTGCTGC TGCTGTATGGGCTGGTGGTGAGCACAGCACTGATCTGGCATCCCATCAACAACTGGCTGCCC TGTTACTGCTGCCCTACCTAGCCTGGCTCACCCTGACTTCAGCCCTCACCTACCACCTGTGGA GGGACAGCCTTTGTCCAGTGCACAGCCTCAGCCACGGAGAAGAGTGACTGAGGCCCTAGGG CATGGGAGAGGAGGGACGCCAGGGTGGGAGGAAGAGTCTGCAAGCAGGGCTGTGGAGTTAG GGTTCACCCCAATGGGACCACCTCCTGGGTCCCCTGGTGCCGTTTTTCCTTA		
	ORF Start: ATG at 36		ORF Stop: TGA at 555
	SEQ ID NO: 234	173 aa	MW at 19491.1kD
NOV63a, CG137707-01 Protein Sequence	MRLQGAI FVLLPHLGPILVWLFTRDHMSGWCEGPRMLSWCPFYKVLLLVQTAIYSVVG YASYL VWKDLGGGLGWPLALPLGLYAVQLTISWTVLVLFVHNPGLYAQLHLHLLLYGLVSTALI WHPINKLAALLLPYLAWLTVTSALTYHLWRDSLCPVHQPPTEKSD		
	SEQ ID NO: 235	624 bp	
NOV63b, CG137707-02 DNA Sequence	AGAGATTGCTCTTCAAGGTGAATGCGGCTTCAAGGGGCTATCTTTGTGCTCTGCCCCACC TGGGGCCCATCCTGGTCTGGCTGTTCACTCGTGATCACATGTCTGGTTGGTGTGAGGGCCCGA GGATGCTGTCCTGGTGCCCATTTCTACAAAGTCTTATTGCTTGTACAGACAGCCATCTACTCTG TCGTGGGCTATGCCTCCTACCTGGTGTGGAAGGACCTGGGAGGGGGCTTGGGGTGGCCCTGG CCCTGCCTCTTGCCCTCTATGCTGTTTACAGCTCACCATCAGCTGGACTGTCCTGGTTCTCTTT TCACAGTCCACAACCTGGTCTGGCCCTGCTGCACCTGCTGCTGCTGTATGGGCTGGTGGTGA GCACAGCACTGATCTGGCATCCCATCAACAACTGGCTGCCCTGTTACTGCTGCCCTACCTAG CCTGGCTCACCCTGACTTCAGCCCTCACCTACCACCTGTGGAGGGACAGCCTTTGTCCAGTGC ACCAGCCTCAGCCACGGAGAAGAGTGACTGAGGCCCTAGGGCATGGGAGAGGAGGACGCC AGGGTGGGAGGAAGAGTCTGCAAGCAGGGCTGTGGAGTTAGGGTTACCCCAATGG		
	ORF Start: ATG at 24		ORF Stop: TGA at 534
	SEQ ID NO: 236	170 aa	MW at 19128.7kD
NOV63b, CG137707-02 Protein Sequence	MRLQGAI FVLLPHLGPILVWLFTRDHMSGWCEGPRMLSWCPFYKVLLLVQTAIYSVVG YASYL VWKDLGGGLGWPLALPLGLYAVQLTISWTVLVLFVHNPGALLHLLHLLLYGLVSTALI WHP INKLAALLLPYLAWLTVTSALTYHLWRDSLCPVHQPPTEKSD		

5

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 63B.

Table 63B. Comparison of NOV63a against NOV63b.		
Protein Sequence	NOV63a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV63b	1..173	130/173 (75%)
	1..170	130/173 (75%)

Further analysis of the NOV63a protein yielded the following properties shown in Table 63C.

Table 63C. Protein Sequence Properties NOV63a	
PSort analysis:	0.6850 probability located in endoplasmic reticulum (membrane); 0.6400 probability located in plasma membrane; 0.4600 probability located in Golgi body; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	Cleavage site between residues 59 and 60

- 5 A search of the NOV63a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 63D.

Table 63D. Geneseq Results for NOV63a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV63a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG89291	Human secreted protein, SEQ ID NO: 411 - Homo sapiens, 170 aa. [WO200142451-A2, 14-JUN-2001]	1..173 1..170	170/173 (98%) 170/173 (98%)	3e-98
AAM80031	Human protein SEQ ID NO 3677 - Homo sapiens, 170 aa. [WO200157190-A2, 09-AUG-2001]	1..173 1..170	170/173 (98%) 170/173 (98%)	3e-98
AAM79047	Human protein SEQ ID NO 1709 - Homo sapiens, 170 aa. [WO200157190-A2, 09-AUG-2001]	1..173 1..170	170/173 (98%) 170/173 (98%)	3e-98
ABB12039	Human benzodiazepine receptor-like protein homologue, SEQ ID NO:2409 - Homo sapiens, 170 aa. [WO200157188-A2, 09-AUG-2001]	1..173 1..170	170/173 (98%) 170/173 (98%)	3e-98
ABG21471	Novel human diagnostic protein #21462 - Homo sapiens, 171 aa. [WO200175067-A2, 11-OCT-2001]	1..173 1..171	125/176 (71%) 134/176 (76%)	3e-61

- 10 In a BLAST search of public sequence databases, the NOV63a protein was found to have homology to the proteins shown in the BLASTP data in Table 63E.

Table 63E. Public BLASTP Results for NOV63a

Protein Accession Number	Protein/Organism/Length	NOV63a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y531	DJ34B21.1 (Novel BZRP (Benzodiazapine receptor (Peripheral) (MBR, PBR, PBKS, IBP, isoquinoline-binding protein)) like protein) - Homo sapiens (Human), 170 aa.	1..173 1..170	170/173 (98%) 170/173 (98%)	9e-98
Q9CRZ8	2510027D20Rik protein - Mus musculus (Mouse), 248 aa (fragment).	1..166 87..248	111/166 (66%) 126/166 (75%)	6e-58
P30535	Peripheral-type benzodiazepine receptor (PBR) (PKBS) (Isoquinoline- binding protein) (IBP) - Bos taurus (Bovine), 169 aa.	8..159 11..158	61/152 (40%) 82/152 (53%)	3e-24
Q96TF6	DJ526114.1 (benzodiazapine receptor (peripheral) (PBR, PKBS, mitochondrial benzodiazepine, MBR) (isoform 1)) - Homo sapiens (Human), 169 aa.	8..159 11..158	63/152 (41%) 83/152 (54%)	1e-23
Q99M32	Similar to benzodiazepine receptor, peripheral - Mus musculus (Mouse), 169 aa.	10..159 13..158	60/150 (40%) 81/150 (54%)	3e-23

PFam analysis predicts that the NOV63a protein contains the domains shown in the Table 63F.

Table 63F. Domain Analysis of NOV63a

Pfam Domain	NOV63a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TspO_MBR	1..157	72/160 (45%) 148/160 (92%)	4.5e-66

5

Example 64.

The NOV64 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 64A.

Table 64A. NOV64 Sequence Analysis

	SEQ ID NO: 237	2560 bp	
NOV64a,	TTGGGAGTCTGAGAAGTCACCACCATGAAGTTGTTCCGGCTTCAGGAGCCGAGGGGCCAGACG GTCCTGGGCTCCATAGACCACCTCTACACGGGTTCCGGGTACCGAATCCGGTACTCGGAAC TG		

CG138033-01 DNA Sequence	CAGAAGATCCACAAGGCAGCTGTCAAGGGCGACGCTGCGGAGATGGAGCGCTGTCTGGCGCGC AGGAGCGGAGACCTTGACGCCCTGGACAAGCAGCACAGGACTGCTCTACATTGGCCTGTGCC AGTGGCCATGTGAAAGTGGTCACTCTCCTGGTTAACAGAAAAAGCCAGATTGATATCTATGAC AAAGAAAAAGAACGCCTTTGATACAGGCTGTCCATTGCCAGGAAGAGGCTTGTGCCGTTATT CTGCTGGAACATGGTGCCAATCCAAACCTTAAGGATATCTACGGCAACACTGCTCTCCATTAT GCTGTGTATAGTGAGAGCACCTCACTGGCAGAAAACTGCTTTTCCATGGTGAATAATTGAA GCACTGGACAAGGACAGTAATACCCCACTTTTATTGCTATAATTTGCAAGAAAGAGAAAAATG GTGGAATTTTATTGAAAAACAAAGCAAGTACACATGCCGTTGATAGGCTGAGAAGAACAGCC CTCATGCTTGCTGTGCACTATGACTCACTGGGTATTGTCAACATCCTTCTTAAGCAAAGTATT AATGTCTTTACTCAAGACATGTGTGGACGAGATGCAGAAGATTACGCTATTTCTTGCCGTTTG ACAAAGATTCAACAACAAATTTTGGAGCATAAAAAGATGATACTTAAAAATGACAAAAATAGAT GTTGGAAGTTCTGATGAATCTGCAGTCAGCATTTTCCATGAAGTGTGTGTGGATTTCATTGTCT GCATTGGATGACGAACCTTTGAGTGTGTGCTAAGCAGTGTGTCCCGAGAAAGTGTGAGAG CCTTTACGTGGACCTTCCCATGGAAGGAAACAGAAATAGTCAATGGAAGGAGAAAGGTCCT CCTGCAAAACATCCTTCTTGAAGCCTAGCACTGAAATGGAAGATCCTGCTGTGAAGGAGCA GTACAAAAAGAAATGTACATGAATTTGTCAACAGAACAGCCTTACCAGTGGCTTCAGAGGAA GAACAGCAAAGGCGTGAAAGAAGTGAAAGAAGCAACACAGGTATATGAAGGAAATATACATA TACAAAAGTGAAAAATACAACATATCAGAAAAATATATGTCATAGTACATTTTCTGCTGCTGT GACAGATTAAACCAACAAAGAAAGATTGGGAAAAACATATCCTCAGCAATTTCCCAAGAACTG AAGGAAGAGCATGATAGGTGCACCTTAAACAAGAAAAATGAAGAAAAACAAATGTTAATATG CTGCACAAAAAAATCGAGAAGAAATAGAAAGGAAAGAGAAACAATATAAGAAAGAAAGTTGAA GCAAAACAACCTTGAAACCACTGTTCATCACTAGAGATGAAACCAAGACTGCAAGAAATACT CCAAATCAGGATTTTCATAATCATGAAGAAGTGAAGATCTGATGGATGAAATGTCATTTTG AAGACAGATATTGCTATACCTCAGACAGGAAATATGCACAAATGAAAAATGACAACTGGAAAAA GAAAAATAATATCTTAAGGACATTAATAATGCTAAAGAAACAAATGCTGCCCTGAAAAGTGT ATAAACTCAATGAGGAAATGATAACAAAAACAGCATTCTGGTATCAACAAGAGCTTAATGAT CTCAAAGCTGAGAATACAAGGCTCAATTCTGAAGTGTGAAGGAAAAAGAAAGCAAGAAAAA CTGGAAGCTGAAATTGAATCTTATCAGTCTAGACTGGCTGCTGCTATAAGTAAACACAGTGAA AATGTGAAACAGAAAGAAACCTAAACTTGCTTTAGAGAGAACACAAGATATTTCTGAGCAA GTAAAAATGAGTTCTGATATTTCCGAAATAGAAGATAAGAAATGAGTTTCTTACTGAACAACTT TCTAAATGCAAAATTAATCAATACCTTAAAGATAAGTTCCGTAAAGACAAGAGATACTCTC AGAAAAAAGTCATTGGCTTAGAACTGTACAAAACGACCTAAGCCAAACACAGCAGCAAATA AAGGAATGAAAGAGATGTATCAAAGTGCAGAAGCTAAAGTCAGTAAATCCACTGGAAGTGG AACTGTGTGGAAGAGAGGATATGTCAACTCCAACGTGAAAATCCGTGGCTTGAACAGCAACTA GTTGATGTTTCATCAGAAAGAGGATCATAAAGAGATAGTAATTAATATCCAAGAGGCTTTATT GAGAGTAGAAAGAAAGACCTCATGCTAGAAGAGAAAAATAGAAAGCTAATGAATGAATATGAT CATTTAAAGAAAGTCTCTTTCAATATGAGAGACAGAAAGCAGAAACAGTAGTAAGTATCAAG GAAGATAAATATTTTCAAACCTTCTAGAAAGAAAGTTAAACATTTGGTTCTGGATACATGTTG AACCTAGTTGAATATAAAAAATCAGTAGGATAAAAAAGTGTG
	ORF Start: ATG at 25 ORF Stop: TAA at 2494
	SEQ ID NO: 238 823 aa MW at 94994.5kD
NOV64a, CG138033-01 Protein Sequence	MKLFGFRSRRGQTVLGSIDHLYTSGSYRIRYSELQKIHKAAVKGDAAEMERCLARRSGDLAL DKQHRTALHLACASGHVKKVTVLLVNRKQCIDIYDENRTPLIQAVHCQEEACAVILLEHGANGP NLKDIYGN TALHYAVYSESTSLAEKLLFHGENIEALDKDSNTPLFAIICKKEKMFVFLKKNK ASTHAVDRLRRTALMLAVHYDSLGI VNILKQ SINVTQDMCGRDAEDYAI SCRLTKIQQQIL EHKKMILKNDKIDVGSSDES AVSIFHELCDVSLALDDELLSVAAKQCVPEKVSPLRGP SHG KGNRI VNGKGEPPAKHPSLKPSTEMEDPAVKGAVQKRM YMNLS TEQALPVASEEEQQRRE RS EKKQPQVYEGNNTYKSEKIQLSENICHSTFSAADRLTQQRKIGKTYPQQFPKKLKEEHDRCT LKQENEEKTNVNM LHKKNREELERKEKQYKKEVEAKQLEPTVQSLEMPKPTARNTPNQDFH NH EEVKDLM DENCILKTDIAILRQEICTMKN DNLEKENKYLKDIKIAKETNAALEKCIKLNEEMI TKTAFWYQQLNDLKAENTRLNSELLEKESKKKLEAEIESYQSR LAAAI SKHSENVKTERNL KLALERTQDISEQVMSSDISEIEDKNEFLTEQLSKMQIKFNTLKD KFRKTRDTLRKKS LALE TVQNDLSQTQQQIKEMKEMYQSAEAKVSKSTGKWN CVEERICQLQRENPNWLEQQLVDVHQED HKEIVINIQRGFIESRKKDLMLEKNRKL MN EYDHLKESLFQYERQKAETVVS IKEDKYFQTS RKKV

Further analysis of the NOV64a protein yielded the following properties shown in Table 64B.

Table 64B. Protein Sequence Properties NOV64a

PSort analysis:	0.7600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV64a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 64C.

5

Table 64C. Geneseq Results for NOV64a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV64a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG08398	Novel human diagnostic protein #8389 - Homo sapiens, 784 aa. [WO200175067-A2, 11-OCT-2001]	370..807 12..449	378/438 (86%) 407/438 (92%)	0.0
ABG04713	Novel human diagnostic protein #4704 - Homo sapiens, 1090 aa. [WO200175067-A2, 11-OCT-2001]	440..807 1..368	319/368 (86%) 345/368 (93%)	0.0
AAE01039	Human death domain-containing receptor (DDCR) protein from HDPDL31 clone - Homo sapiens, 281 aa. [WO200129063-A2, 26-APR-2001]	181..461 1..281	236/281 (83%) 252/281 (88%)	e-129
ABG01862	Novel human diagnostic protein #1853 - Homo sapiens, 307 aa. [WO200175067-A2, 11-OCT-2001]	111..385 55..301	178/297 (59%) 192/297 (63%)	8e-84
ABG04687	Novel human diagnostic protein #4678 - Homo sapiens, 418 aa. [WO200175067-A2, 11-OCT-2001]	3..164 130..291	142/162 (87%) 148/162 (90%)	9e-77

In a BLAST search of public sequence databases, the NOV64a protein was found to have homology to the proteins shown in the BLASTP data in Table 64D.

Table 64D. Public BLASTP Results for NOV64a				
Protein Accession Number	Protein/Organism/Length	NOV64a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H0H6	Hypothetical 94.0 kDa protein - Homo sapiens (Human), 823 aa.	1..823 1..823	717/823 (87%) 762/823 (92%)	0.0

CAB99338	BA255A11.3 (novel protein similar to KIAA1074) - Homo sapiens (Human), 1117 aa.	2..796 3..782	279/820 (34%) 420/820 (51%)	e-103
Q9H560	BA526D8.2 (novel protein similar to KIAA1074) - Homo sapiens (Human), 264 aa.	2..261 3..262	169/260 (65%) 212/260 (81%)	1e-94
Q9HIQ1	BA145E8.1 (KIAA1074) - Homo sapiens (Human), 1710 aa.	431..812 780..1223	195/444 (43%) 275/444 (61%)	8e-89
Q9UPS8	KIAA1074 protein - Homo sapiens (Human), 1709 aa.	431..812 779..1222	195/444 (43%) 275/444 (61%)	8e-89

PFam analysis predicts that the NOV64a protein contains the domains shown in the Table 64E.

Table 64E. Domain Analysis of NOV64a			
Pfam Domain	NOV64a Match Region	Identities/ Similarities for the Matched Region	Expect Value
ank	66..98	14/33 (42%) 28/33 (85%)	1.9e-07
ank	99..131	12/33 (36%) 25/33 (76%)	1.2e-05
ank	132..164	11/33 (33%) 27/33 (82%)	7.9e-06
ank	165..197	11/33 (33%) 26/33 (79%)	0.00014
ank	198..230	11/33 (33%) 23/33 (70%)	0.0032

5

Example 65.

The NOV65 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 65A.

Table 65A. NOV65 Sequence Analysis			
	SEQ ID NO: 239	14994 bp	
NOV65a, CG138043-01 DNA Sequence	GCTTCTCGACTGGGACGCGGCGGAGGAGGAGCGCGGCGGCCCGAGTCTCCCTGAGCCATG GGCAACGAGGCGAGCTTGAAGGGGAAGGGCTCCCGAAGGGCTGGCGGCGGCCGAGCGGCT GGAGGAGGAGCTAGCGGGGCGGGGAGCCCTCTCACACCGCGATCCCGGCCGCGATGGAGGCG GATTTGAGCCAGCTGAGCGAAGAGGAGAGGAGACAGATCGCCGCTGTCATGTCAAGGGCGCAG GGGCTGCCCAAGGAAGCGTCCCCCGGCCGCTGCGGAGTCGCCTTCATGCACAGGAACAA GAGTTGGATAGTAGTCATCCTCAAAGCAATCAGGAAGACCCCGGACCCCTGGGCGTCCAGCT CAACCTGGTCTCAGTAAAGTAGAACTACAGACACTTTCAGGTCAGAGCAGAAATTGCCTGGG		

AGGAGTCCTTCCACTATTAGCTTGAAAGAATCAAAGTCCAGAACTGATTTAAAGGAAGAGCAC
AAGTCTAGTATGATGCCTGGCTTCTCTCAGAGGTAAACGCTTTAAGTGTGTTTCTCTGT
GTAAATAAATTCAACCCTTTTGATTTGATATCAGACTCTGAGGCATCCAGGAAGAAACCACC
AAGAAACAAAAGTGGTTTCAAGGAGCAAGGAAAACCTGAAGGAATCATAAAACCTCCTTTA
CAACAACAGCCACCCCAAGCCGATTCCTAAGCAGCAAGGACCTGGTAGGGATCCGCTTCAGCAG
GATGGCACTCCCAAATCAATATCTTCTCAACAACAGAAAAAATTAAATCACAACTCCAGGT
ACAGGAAAGCCAATTTCAGGGTCTTACCAGACTCTCAGACAGACCATGCAAAATTGCCACTT
CAACGAGATGCATCCAGGCCTCAGACTAACAGGCAGACATAGTAAGGGGAGAATCAGTTAAA
CCCTCACTGCCAAGCCCATCCAAACCACTATTTCAGCAACCAACTCCTGGAAAACCTCCAGCA
CAGCAGCCTGGACATGAAAAATCACAGCCTGGGCTGCAAAAGCCCCAGCTCAGCCCTCAGGG
CTAACAAAGCCATTGGCTCAACAACAGGGACAGTGAACCCCCAGTCCAGCCACCAGGGACA
ACAAAGCCTCCAGCTCAGCCTCTTGGTCTGTAAAGCCTCCAGCTCAGCAGACTGGGTTCAGAG
AAGCCTTCATCGGAGCAGCCTGGGCCAAAGGCTTTAGCTCAGCCTCCTGGAGTTGGAAAGACT
CCAGCTCAACAGCCAGGGCCAGCAAGCCTTCAACCCAGCAGGTGGGGACACCAAAACCCCTA
GCTCAACAACCTGGGCTACAGTCTCCAGCTAAGGCACCTGGGCTTACAAAGACTCCAGCTCAG
ACAAAGCCCCATCTCAACAGCCTGGCTCAACAAAACCCCCACCTCAACAGCCTGGCCAGCA
AAGCCCTCACCTCAACAGCCTGGCTCAACAAAACCCCCATCTCAACAGCCTGGCTCAGCAAAA
CCCTCAGCTCAACAGCCTAGCCAGCAAGCCCTCGGCTCAGCAATTTACAAAACAGTAAGC
CAACAGGATTTGGAAAACCTCTGCAGCCACCAACAGTGTCTCCATCTGCAAAAACAGCCTCT
TCACAAGGCCTCCCTAAAACCATCTGTCTCTTTGCAATACCACTGAACCTCTGTTGCATGTT
CCAGAAAAGGCCAATTTAACACATGCACTGAGTGTCAAACCACTGTCTGTAGTCTCTGTGGT
TTTAAATCCCAATCCTCATTTAACGGAGGCAAAAGAGTGGCTCTGTTTGAAGTGTCAAATGAAA
AGAGCTCTAGGCGGGGATCTGGCTCCAGTTCCTCATCACCCAGCCCAAACTGAAGACTGCA
CCTGTTACCACTACATCAGCAGTGAGCAAAATCATCCCCACAGCCACAGCAGACTTCCCCAAAG
AAGGATGCTGCACCAAAACAGGATCTCTCAAGGCACCTGAGCCTAAAAAGCCACCACCGCTA
GTGAAACAACCAACCTTTCATGGCTCTCCTTCAGCCAAAGGCCAAGCAGCCCCCTGAGGCAGAT
TCTTTGTCCAAGCCAGCCCCCTCCAAAGAACCTTCTGTCCATCTGAGCAGGACAAGGCCCT
GTTGCTGATGATAAAACCAAGCAGCCCAAGATGGTAAAGCCAAACCACTGACCTTGTATCTCA
TCATCAGCAACAACAAAACCTGATATTCCAAGCTCCAAAGTACAGTCACAAGCTGAAGAGAAA
ACAAACCCCTCCTCTAAAACAGACTCTGCCAAACCTCAGAGATTTTCCACCAACAGGGGAA
AAAGTCACCCCATTTGATTCTAAAGCCATACCTCGACCTGCATCAGATTTAAAAATTATTTCA
CATCCTGGTCCAGTTCAGAGAGCAAAAGGTCAAAAACAAGTTGACCCCTCAAAAAGAGGAA
GAACCAAGAAAGCACAACCAAAATGAGTCTTAAACAGATGCCAAGCCAATGCCAAAAGGG
TCACCAACACCCCTGGCCACGACCTACCGCTGGCCAACTGTCCCCACACCTCAACAGTCC
CCAAAGCCTCAGGAGCAGTCAAGGCGTTTCAGTCTGAATCTGGGAAGTATTACTGATGCCCCC
AAATCACAGCCTACAACCTCTCAAGAGACCGTGAAGTGGGAACTCTTTGGGTTTGGAGCATCA
ATCTTCAGCCAGGCATCAAAATTTAATTTCCACTGCAGGCCAACCTGGACCTCATTCACAAAGT
GGACCAGGGGCCCAATGAACAAGCCCCCTGCCCTTCACAGCCACCTACTTCACAAGGGCCA
CCCCAATCCACAGGTCAAGCAACCAAGCAGCCTGCAAAAAGTATACCTGTGAAAAGGAAACA
AAAGCCCCAGCAGCTGAAAAATTAGAGCCCAAGCTGAACAAGCTCAACAGTAAAAAGAAACA
GAAACAGAAAAAAGCCACCTATTAAAGGATAGCAAACTTTAAACAGCTGAGCCTCAAAAG
GCTGTCTTCCACAAAACCTGGAGAAATCGCCCAACCAAGATCAACCTGTCTCTCTGCAAA
ACTGAACCTCAACATAGGTTCTAAGGATCTCTTAACCTCAATACTTGCAGTGAATGCAAGAA
CAAGTGTGTAATCTCTGTGGATTAAACCTACACCACATTTGACTGAGATTCAAGAATGGCTT
TGTTAAATTGCCAAACCCAGAGAGCAATATCAGGACAGCTTGGAGACATACGCAAAATGCCA
CCTGCACCATCAGGACCCCAAGCATCTCTATGCCTGTTCTTACAGAAATCATCTCAGAAA
ACAGCAGTGCCTCCCAAGTAAATTAGTGAAGAAGCAAGAACAAGAATAAAAACGGAAGCT
GAAAAAGTCATTCTGAAAAAGTAAAGGAAACACTATCAATGAAAAAATTCCTCCTATGGTA
ACCACAGATCAAAAACAAGAAGAGAGTAACTAGAGAAAGACAAAGCTTCAGCTCTTCAAGAA
AAAAAGCCACTCCCTGAAGAAAAAATAATCCCTGAAGAAAGAAAGATACGTTCTGAAGAA
AAAAAGCCACTCTAGAAGAAAAAAGCCAAACCCCTGAAGACAAAAAGCTACTCCAGAGGCA
AAAACATCAGCCCCAGAAGACAGAAACATGACTTACTTAAATCTCAAGTACAAATTGTGAA
GAAAAGCTTGAAGGCAGAGTGGCTCCAAAGCAGTGCAAGAGGGAAACAACCAACAGACCAAG
ATGGAAGGTTTACCATCTGGCACACCTCAGAGTTTACCTAAAGAAGATGATAAGACAACCAAA
ACAATAAAGAAGACGCCACGCCACCATGCACAGCAAAACCTGATCAGGAAAAGGAAGATGAC
AAATCAGACACCTCAAGTTCTCAGCAGCCTAAAAGCCCCAAGGTCTGAGCGACACGGGATAT
TCTTCCGATGGAATATCAAGCTCAGTTGGTGAATTTCAAGTCTTATTTCAACTGATGAAAAG
GATATTCTCAAGGGACTCAAAAAGGACTCTTTTTCACAAGAAAGCAGCCCTTCCAGCCCTCA
GATTTGGCTAAGTTAGAAAGTACAGTCTCTATCTATTTTGAAGCTCAAGCAAGTACACTGTG
GATGAAAAGTCAGAAAAGAAAAACAACCCCATGAAGTTTCTCTGAACAGCCTAAAGACCA
GAGAAAACCTCAGAGTTTATCTGAACCTTGGAAATTAATTTTCAAGAGAGGAGATCAAGAG
AGTCAAGAAGAAAGGAAAGACACTTTTAAAAAGATAGCCAAAGATATTCCTTCCAGCAAG
GACCATAAAGAGAGTCTGAGTTTGTGATGACATAACTACTAGAAGAGAGCCTTATGATTCA
GTTGAAGAGAGATGAAAGTGAAGTCAAGTCAAGTGTTCACAAAGAAAGCAAGCAAGTGTGTT
GGCTCATCAAGCAGTGATGAGTATAACAGGAAGACAGCCAAAGGATCAGGGGAGAGGAGGAC
TTCATTGAAAAACCAATCATAGAAATGAGTGTGATGAAGATGCTTCAGGTTCTGAAGATGAT

GAGTTCATCAGAAACCAGCTCAAAGAGATTAGTAGCAGTACTGAGAGCCAGAAGAAGGAAGAA
ACAAAGGGAAAAGGCAAAATAACAGCAGGGGAAACACAGACGACTGACTCGAAAAAGTAGCACA
AGCATTGATGAAGATGCAGGAAGACGTCATCATGGCATGATGAAGACGATGAAGCATTGAT
GAAAGTCCTGAACCTAAATACAGAGAACTAAAAGTCAGGAAAGTGAAGAACTTGATGTTACT
GGAGGAGGAGGGCTACGCCGATTTAAAACAATTGAGCTCAACAGTACAATAGCAGATAAATAT
TCTGCAGAGTCATCACAGAAAAAACAAGTTGTATTTTGACGAAGAGCCAGAATTGGAAATG
GAAAGCCTGACAGACTCACCTGAAGATAGGTCAAGGGGAGAGGGATCTTCGAGTCTGCATGCT
TCCAGCTTCACTCCTGGTACATCCCTACATCAGTATCATCACTTGATGAGGACAGTGACAGT
AGCCCGAGTCACAAAAAAGGAGAGAGCAACAGCAACGCAAGCTCGGCACAGACCACATGGC
CCTCTTTTGCTTACTATTGAAGATTCTTCAGAGGAAGAAGATTGAGAGAGGAAGAAGATTA
TTAAAGGAGCAAGAAAAGCAGAGGGAAATAGAACAGCAACAAAGAAAGAGTTCTAGTAAAAAA
TCAAAGAAAGACAAAGATGAACCTCGAGCTCAGAGAAGAAGGGAAAGGCCAAAGACACCACCT
AGTAATCTCTCTCCATTGAAGATGCATCTCCGACAGAAGAGTTACGTCAGGCTGCAGAAATG
GAGGAGCTCCATAGATCTTCTGTTCTGAATATTCACCTAGCATAGAGTCCAGCCAGAAGGT
TTTGAATAAGCCCGGAAAAAATAATAGAAGTACAAAAAGTTATAAATTGCCACAGCTGTT
TCATTATACTACCAACAGATGAGCAATCTATTATGCAGAAAGAAGGTAGCCAAAAGGCGTTA
AAAAGTCTGAGGAGATGTATGAAGAAATGATGCATAAAACACACAAATACAAAGCTTTTCCA
GCTGCAATGAACGAGATGAAGTGTGAAAAAGAGCCTTTGTATGGTGGGATGCTAATAGAG
GATTATATTTATGAATCTTTAGTAGAAGACACGTACAATGGATCGGTAGATGGCAGTCTGCTA
ACAAGGCAAGAAGAAGAAAATGGATTTATGCAGCAGAAAGGAAGAGAGCAAAAGATAAGACTT
TCAGAACAGATTTATGAAGATCCTATGCAGAAAATTACAGACCTCCAGAAAGAGTTTATGAG
TTAGAAAGCTTACATTCTGTTGTGCTCAGGAAGATATTGTTCAAGCTCTTTATCATCCCA
GAAAGCCATGAGATAGTGGACCTGGGTACTATGGTAACTTCTACAGAAGAAGAAAGAACTA
CTAGATGCTGATGCTGCCATATGAAGAACTTATGAAGAGGCAACAGATGCAATTAAACCTGGA
TCTAGCCCAACCCAGGCCCCCATTTGGTGAGGATATGACAGAGTCCACCATGGACTTTGACAGA
ATGCCAGATGCCTCTTTGACATCAAGTGTCTCTCAGGAGCGTCTCTTACAGATTGACCCAGC
AGTGCAACACTCTCTATCCAGATGTTAAAATAACCCAACATTTTCAACAGAAGAAATGAG
GATGAATATGTAACCGATCATACAAGAGAAATTCAGAGATAATTGCCATGAATCGCTGATT
TTGACCTACTCGGAGCCTTCAGAAAGTGCTACATCTGTCCCACTCTGACACACCTTCTCTC
ACATCATCTGTTTCTTCCGTCTGTACCACAGATAGCTCTTACCCATTACTACCCTGGATAGC
ATAACCACAGTTTATACAGAGCCAGTGGACATGATAACTAAATTTGAAGATTCTGAGGAAAT
TCTTCATCAACTTATTTCCAGGCAGCATTATAGACTATCCAGAAGAAATAAGTGCACTTTA
GATCGGACTGCCCAACAGATGGTAGAGCTAGTGCTGATCATATTGTTATTTCTTATCTGAT
ATGGCATCTTCTATCATAGAATCTGTAGTACCTAAACCTGAAGGGCCAGTTGCTGACACTGTT
TCTACTGACTTACTTATATCTGAAAAGGACCCAGTGAAGAAAGCCAGAAGGAAACTGGGAAT
GGAATCATTTGGAAGTTTGGAAAGCTTACAGAGATAAAAGGAGTTGGAGGCCGAACGAACA
AAAAGTAGCTTATCCGAAACCGTGTGATCACCCACCTTCTTCTGTAATAGCCCTTCCAATG
AAAGAGCAGCTTTCAACTACATACTTTACATCTGGAGAGACCTTTGGTCAGGAAAAACCTGCA
TCTCAGTTACCATCTGGCAGTCTTCTGTTTCTCTCTTCCAGCTAAACCTCGCCCATCTTT
AGAAGTTCTTCTTGGATATATCAGCTCAACCTCTCTCCCTCTCTCCCTCTCTCTCTCTCT
CCTCTCCACCACCACCCCTCTCTCCCAACCTCTCTCCCAACCTCTCTCAACCACTCTCAACCACT
ATTCTTCTTAAAAAAGTTAACAGTTGCATCTCCAGTACTACAGCTACACCTCTGTTTATGAT
GCTGTTACTACTCTAGAGACCACAGCTGTTCTGAGAAGTAATGGATTACCTGTTACAAGAATA
TGTACTACTGCACCTCTCTGTTCTCTTAAGCCATCTTCAATTCCATCTGGACTTGATTTT
ACCCACAGGCTGAGCCAAGCAACCTCCAATCGCCCCAAACAGTGATTTCTCAGCTTCCA
ACAACCTACAAAAAACCACAGATATACACCCCAACCAACAGGCTATCTTTAACTTCAAGT
ATGACCTTAAATTCAGTGACTTCAGCAGATTATAAATGTCCTTCCCTACCTCCCACTTTCC
CCACACTCCAACAAGTCTTCAACCAAGATTTTCCAATCCCTCACAGAACTTATGTAGTTATT
ACATTGCCATCTGAACCAAGGACTCCAACAGATTCTTCTGCTAGTCAAGCAATTACCAGTTGG
CCCTTGGGATCACCTCCAAGATCTGGTTCTGTTGAACCTGTGTTTCTGCTGATGTAATTGATTG
GTGACAGCTGTAGAAATCCAATTTCTTCAGAACAGACCTTCTACATCTCTGGAGCTTTACAG
ACATTTTCTGCTACCCCTGTACAGCACCTCTTCAATTTCAAGCAGCTCCACATCAGTTACA
CAGTTTCTCACTACTGAAGTTTCCAAGACTGAGGTTTCAAGCAACCAAGATCAGCTCTAGT
GTTGGTCTCAGCAGCATTCCATAACAATTCCTCCAGAGCTCTTGTCTAGATAACATACAT
TTAGAGAAGCCTCAGTATAAAGAGATGGAAAAATGCAACTTGTGTTGATGTAATTGATTG
CGTACAGTACCAAGGTAGAAGTTAAAACAACGTATAAATGTATTGATCTTCTGCTTCTACA
ATGGATGTGAAAAGGCAGATCACAGCAATGAAGTTTATGGGAAACAAATAGTGTGTCCAA
CCCTCTATTATAAATCTTAGTGACATCATCAATAGTACTCTGTATCTCTGGCCACTGAG
ACAGTGACCTTTGTACATGCACAGCTAGTGCAAGTTACACTACAGGCACAGAAAGCCTAGG
GGTGCAACATGCAATGACAACCACTCCAACCTACAACATCAAAGCATGCTGAGCCCCCA
TACAGGATACCAAGTGACCAGGTCTTCTCTATAGCTAGGGAAGAAGCAACATAAATCTATCT
CTAGGTACTCCAGCATGCAGTGACATTGGCTATTACAAACCTGTCAGTGTGCTCTGTT
GGTGTCAAAATGGATGGACTGATAGCACCGTATCCAGGGAATCACTGATGGGGAAGTAGTG
GATCTCAGTACAACCAAGTCTCACAGAACAGTCGTAACCAATGGATGAGTCTACTTCAAGTGTG
ATGACCAAAATAATAGAAGATGAAAAACCGTTGATTAAACCGAGGAGAAGAGCTGTGTGC
TGTGATGTGGTTATAAATTACCATTTGGAAGGAGCTGCACAGCAGCAGCTGCAACTACT

CTTCCTGAGGATCGTTTTGGTTATAGGGATGACCACTATCAGTATGATCGATCAGGGCCATAT
GGTTATAGAGGGATTGGGGGAATGAAGCCTTCCATGTCTGACACAAATTTAGCAGAAGCTGGA
CATTTTTTCTATAAAAGTAAGAATGCTTTTGATTATTCTGAAGGAAGTACACAGCAGTAGAT
CTGACTTCAGGGAGAGTTACTACAGGTGAGGTAATGGATTATTCAAGCAAGACTACAGGTCCA
TATCCAGAAACACGACAAGTCATTTAGGAGCTGGGATTAGTACCCACAGTATTCCACAGCA
AGAATGACACCACCAGGACCCAGTATTGTGTGGGAGTGTTTTGAGGTCTATCTAATGGT
GTTGTCTATTCTTCAGTAGCAACTCCAACACCCTCTACATTTGCTATCACCACACAACCTGGC
TCCATTTTCAGCACCACAGTGAGGGATTGTCTGGTATTCATACGGCTGATGCAGTGACTTCA
TTACCTGCCATGCACCATAGCCAGCCAATGCCCTAGATCATATTTATAACAACAGGTGCATCT
GAAACGGACATTGCAGTAAGTGGTATTGATATCAGTGCCAGTTTGCAAATATTACTATGGAG
TCTCTTACTGCTGAGACGATAGACTCTGTTCCCACTTTAACCACAGCATCCGAAGTGTTCCT
GAAGTGGTGGGAGATGAAAGTGCTCTTTAATGTCCCTGAAGAAGATAAACAACAGCAGCAG
CTAGACTTGGAGCGTGAGCTCCTGGAAGTGGAGAAAATTAAGCAACAGCGCTTTGCTGAGGAA
TTGGAGTGGGAACGTGAGGAATTCAGAGTTCCGAGAACAGAAAGATCATGGTTTCAGAAA
AAGTTGGAGGAGCTGCAGTCTATGAAGCAACACCTTCTCTTTTCAGCAAGAAAGAGCGGCAA
GCCCAGTTTCATGATGAGGCAGGAGACGTTAGCTCAGCAACAGTTACAGCTTGAGCAGATCCAA
CAGCTGCAACACAGCTTACCAGCAGCTGGAGGAGCAAAAGATTTCGGCAGATCTACCAGTAT
AACTATGACCCTTCTGGAAGTCTTCCACAAACCACTACAGAGCAGGCAATTTTGAAGGT
CAGTATGCTGCTCTGGAAGGCAGCAATTTTGGGCAACTGAAGATGCAACCACACAGCTTCA
GCTGTTGTGGCAATTGAAATACCACAAAGCCAAGGATGGTACACCGTTTCAGTCTGATGGTGT
ACTCAGTACATTGCCCCACCTGGTATCCTGAGCACTGTTTCAGAAATACCTCTAACAGATGTT
GTTGTGAAGAGGAAAAACAACCCAAAAAGAGAAGTTCTGGAGCTAAAGTCCGAGGACAGTAT
GATGACATGGGAGAAAATATGACAGATGATCCCCAAGTTTAAAAAGATAGTGGACAGTGGT
GTACAAACGGATGACGAAGATGCCACAGATCGGAGCTATGTGAGTAGGAGAAGGAGAACTAAA
AAGAGTGTGGATACAAGCTCCAACTGATGATGAAGATCAGGATGAGTGGGATATGCCTACT
AGATCAAGGAGGAAAGCTCGTGTAGGGAAATATGGTGACAGCATGACAGAGGCTGACAAGACC
AAACCCCTTTCCAAAGTCTCCAGCATAGCAGTTCAACCGGTAGCAGAGATATCTGTGCAAACT
GAACCCAGTTGGAACCATAGAAACACCCTCCATACGGGCACGAGTGGATGCCAAGGTAGAAAATA
ATTAAACACATTTTCAGCACCTGAAAGACTTACAAAGGGGGCAGTTTAGGATGTCAAACAGAA
GCAGATTGACACACAAAGTCTCTCAATATCTGAGTGCCACATCTCCACCCAAAGACAAGAAA
CGCCCAACACCTTTAGAGATTGGTTATTCATCTCACCTCCGGGCAGATTCCACAGTACAGCTG
GCTCTTTCCCAACCAATCCCAAGTCTTTACTCACCATCTCACCACTTTCACCAAGC
AAAGCCTTAGAATCAGCCTTTGTACCTTATGAAAAACCCCTCCCTGATGATATAAGTCCACAG
AAAGTACTGCATCCAGATATGGCTAAAGTTCCCCCAGCAAGTCTTAAGACAGCCAAGATGATG
CAGCGTTCTATGTCTGACCCCAAGCCTCTGAGTCCACAGCAGACGAAAGTTCCAGGGCTCCT
TTTCAGTATACCGAGGGCTATACGACTAAAGTTCTCAAACCATGACATCTCTGAGGCCAG
AAAAAAGTTAAAAAGAACTCTGCCAAATCCACCTCCTGAGGAGATTTCACAGGAATCAATCC
ACATTGACACAATGGGCACAGTTTCCAGGAGAAGGATCTGCAGAACCAACACAATGGCACGA
GCCAAGATTCTCCAGGACATAGACAGAGAGCTTGATCTTGTGGAAAGGGAGTCTGCAAACTT
CGAAAGAAACAAGCAGAGCTTGATGAAGAAGAAAAGGAGATTGATGCTAAGCTACGATACCTG
GAAATGGGAATTAAACAGGAGGAAAGAGGCCCTATTAAAGGAGAGAGAAAAGACGAGCC
TACCTCCAGGGAGTAGCTGAGGATCGTGATTACATGTCTGACAGTGAAGTGAAGTACACAAGA
CCAACCCGAATAGAAAGTCAGCATGGCATTGAGCGACCAAGAACTGCTCCCCAACTGAATTC
AGCCAGTTTATACCACCACAACCCAAACAGAATCTCAACTAGTTCTCTCCGACAAGTCTTAC
ACACAATACCAGTACTCTTCCCTGCTCTCTACCCAAGCACCACCTCATACACTCAACAG
TCTCATTTTGAGCAACAACTTTGTACCATCAGCAAGTTTACCTTATCAGACTCAGCCAACA
TTCCAAGCTGTGGCAACAATGTCTTCCACCTCAAGTTCAACCTACACCAACCCACAGCCT
TCTTATCAGTTACCTTCAGAGATGATGGTGATACACAGAAGCCACGGCAAACTACATTATAT
TTGGAGCCCAAGATAAATCTCAAATATGAAGTGATTGCAACCAACCCCTTATGATAGCACCT
GTTTCTACGGATAACACATTTGCTGTTTCCCATCTTGGTAGTAAGTACAATAGTTTAGACTTG
AGAATAGGTTTGGAGGAAAGAGTAGCATGGCAAGCAGTCCAATATCAAGCATATCTGCAGAT
TCTTCTATGCAGATATTGATCACCATACTCCACGAAATTATGTCTTAATTGACGACATTGGA
GAGATCACCAAAGGAACAGCGGCATTAAGCACCAGCATTAGCCTTCATGAAAGGATCTGTCA
AAAACAGACCGTCTCTTCCGAACCACTGAGACACGCCGTCTCAAGAAGTGACAGATTTCTTA
GCACCTTTACAGTCTTCTCTAGATTGCATAGTTATGTGAAGCGGAGGAAGACCCCAATGGAG
GATCCTTACGAGTTAAAGCTTCTGAAACATCAGATTAAACAGGAATTTCTGAGAGGGACAGAG
AGCTTAGATCACCTTGCTGGTCTTCTCATTATTACCATGCTGATAGCTACAGACATTTT
CCAAAATCTGAGAAGTATAGCATCAGTAGACTCACACTTGAAAAACAAGCAGCAAAACAAGT
CCAGCAGCCATACTTTATCAAAAGCAGTCAAGCATAAGAAATCACTAATTGACCCCAATGATG
TCAAAATTTTACCTATTCAAGAAAGTAGAGACCTTGAACCTGATTATTCAAGCTATATGACT
TCTAGCACTTCATCTATTGGTGGCATTCTCTCCAGGGCAAGGCTCCTTCAAGATGACATCACT
TTTGGCCTCAGAAAAATATTACAGACCAACAAAATTTATGGGATCTTCTTGGCACAGGA
CTGGGCACATTAGGAAATACCATACGCTCAGCTCTGCAGGATGAAGCGGATAAGCCATACAGT
AGTGGCAGCAGGTCCAGACCTTCTCTCAGACCTTCTCTGCTATGAGGCTTGAATTTATCAAT
AAAAGGGATTCTTCTAGCTCTTCCCTAAGACTGAAAGCTCAAGAGGCTGAAGCTCTAGATGTT
TCCTTTAGTCATGCATCATCTCTGCGCAAGTAAGCCGACCAAGTTTGCCAATTAGTCAAAGT

	<p>AGAGGAAGAATACCAATTGTGGCCGAGAATTCTGAAGAAGAAAGCCCACTCAGTCCTGTTGGC CAGCCAATGGGAATGGCCAGGGCTGCAGCTGGACCCCTGCCACCAATATCTGCAGACACCAGG GATCAGTTTGGATCAAGCCACTCATTGCCTGAAGTTCAGCAACACATGAGGGAAGAAATCACGG ACTCGAGGCTATGACCGTGACATAGCATTATCATGGATGACTTCCAACATGCCATGTTCAGAC AGTGAAGCCTATCATCTGCGTCGTGAGGAAACAGATTGGTTTGATAAACCCAGGGAGTCTCGT TTGGAAAATGGACATGGTCTGGACCGAAAACCTGCCGAAAGATTGGTCCACTCTAGACCACCTC AGTCAACATCAAGAGCAAATTATACAGATGAACGGGAAAACCTATGCACTACATCTTCTCTCAC GCAAGGATAAAAATAACAAGAGACTCAAAGGATCACACAGTTTTCAGGTAATGGATTAGGAATT AGAATTGTGGGTGGTAAAGAAATCCCGGGACATAGTGAGAAATTGGAGCCTATATTGGCAAG ATTCTTCTGGGGGAAGTGCGGAACAGACGGGGAAGCTTATGGAAGGGATGCAAGTATTGGAA TGGAAATGGAATCCCTTGACTTCTAAACATATGAAGAAGTTCAGAGTATCATTAGTCAGCAA AGTGGGGAAGCAGAAATATGTGTAAGACTGGACCTCAATATGCTATCAGATTCTGAAAATTCC CAGCATCTGGAACCTCATGAGCCACCAAAAGCTGTGGATAAGGGGAAATCCCAGGGGTTGAT CCTAAGCAGTTGGCAGCAGAACTCCAGAAGGTTTCACTACAGCAGTACCCTGGTCTGTCTCA TCAGTTGTTGAAAAAGGATCTCATGTTTATTGAGGTCCTACATCAGCAGGATCCAGTTCCGTT CCCAGCCCTGGGCAACCAGGGTCCCCCTCAGTGAGCAAAAAGAACGACGGCAGCAGCAAGCCT ACCGATGGAACAAAGGTTGTCTCTCATCCAATTACAGGAGAAATTCAGCTTCAAATTAACAT GATCTTGGAAATCTCATAATACATATTCTCCAAGCAAGAAATCTTGTCTCTGAGACAACAAAT GGTTATTCTGACCCTTTTGTGAAAGTGACCTTCTTCCAGGGAGAGGTCAAGTCATGGTTGTC CAGAAATGCAAGTGCTGAGTACAAGAGAAGGACTAAACATGTCCAGAAAAGTCTTAATCTGAG TGGAAATCAAACAGTAATTTATAAAAGTATTTCCATGGAACAGCTCAAGAAGAAAACACTGGAG GTGACAGTTTGGGATTATGATAGATTTTATCCAACGACTTCTTGGGGAGGTATTGATTGAT TTATCTAGCACATCTCACCTGATAACACTCCAAGGTGGTATCTCTCAAAGAACAGACTGAA AGCATTGATCATGGCAAGTCTCATTCCAGTCAGAGCAGCCAGCAGTCCCCAAAGCCATCTGTT ATCAAAGCAGAAGCCATGGTATCTTCCCTGACCCATCAAAGGACATGCAGGTTCCACCATT GAGAAATCCCATAGTAGTCTGGTAGCTCAAATCATCATCAGAAGGCCATCTCCGTTCTCAT GGACCATCTCGCAGTCAAAGCAAAACCAGCGTCACTCAGACCCACCTGGAAGATGCAGGGGCT GCCATAGCTGCTGCCGAAGCTGCCGTGCAACAACCTCCGATTCAACCAAGTAAAGACGCAAA TAAATTCCTCAGCATGGCAGCTTAATGTTTATCTGTTGCTTTCTTTCTGCTGCTCTTCTCT GTTTGCTTTTCAATCTCTCTGCTCACCCTGTTCTCTGTCCTTTGTCTGTGTAAGA ACGATATAAATACATTAATATGCTCTTCTTATTATTAGATTTTTTTTATTACATAGACTGAAA TAAACTGGCTGTTTCTCTTGTGTTTCCACCATCCATCAACCTGGCTCATGACATTGTATAC AGTGTCTGTGATGTTTGAAGCAAGCAATGTTGTGTCTCTTTTGTGCTTAAATATC</p>
	<p>ORF Start: ATG at 61 ORF Stop: TAA at 14680</p>
	<p>SEQ ID NO: 240 4873 aa MW at 531766.4kD</p>
NOV65a, CG138043-01 Protein Sequence	<p>MGNEASLEGEGLPEGLAAAAAGGGASGAGSPSHTAIPAGMEADLSQLSEERRQIAAVMSRA QGLPKGSVPPAAAESPSMHRKQELDSSHPPKQSGRPDPGRPAQPGLSKSRRTDTFRSEQKLP GRSPSTISLKESKSRDLKEEHKSSMMPGFLSEVNALSAVSSVVKFNPFDLISDSEASQEET TKKQKVQKEQKPEGIIKPPLQQPPKPIPKQQGPGRDPLQDGTPKSISSQPEKIKSQPP GTGKPIQGPQTPTQDHAHLPLQRDASRPQKQADIVRGESVKPSLPSPSKPPIQPPTPGKPP AQQPGHEKSQPGPAKPPAQPGLTKPLAQPGTVKPPVQPPGTTKPPAQLGPAKPPAQQTGS EKPSSEQPGPKALAQPPGVGKTPAQPPGPAKPPTQVGTGPKPLAQPGQLQSPAKAPGPTKPA QTKPPSQPGSTKPPQPGPAKPSPPQPGSTKPPSQPGSAKPSAQPPSAKPSAQQFTKPV SQTGFGKLPQPPVSPSAKPPSQGLPKTICPLCNTTELLHVPEKANFNTCTECQTTVCSLC GFNPNPHLTEAKEWLCNLCQMKRALGGDLAPVSSPQPKLTAAPVTTTSAVSKSSPPQQTSP KKDAAPKQDLSKAPKPPPLVQPTLHGSPSAKAKOPPEADSLKPAKPPKPSVPSEQDKA PVADDKPKQPKMVKPTTDLVSSSSATTKPDIPSSKVQSQAEEKTPPLKTDKSAKPSQSFPTG EKVTPFDSKAIIPRASDSKIIISHPGPSSESKGQKQVDPVQKKEPKAQTMSKPKDAKMPK GSPTPPGPRPTAGQTVPTPQQSPKQEQSRRFSLNLGSIIDAPKSQPTTPQETVTGKLFGFGA SIFSQASNLISTAGQPGPHSQSGGAPMKQAPAPSQPTTSQGGPKSTGQAPPAPAKSI PVKKE TKAPAAEKLEPKAEQAPT VKRTETEKPPPIKDSKSLTAEPQKAVLPTKLEKSPKPESTCPLC KTELNIGSKDPNNTCTECKNQVCNLCGFNPTPHLTIQEWLCLNCQTQRAISQGLDIRKM PPAPSGPKASMPVPPTESSQKTAVPPQVQLVKKQEVEVTEAEKVILEKVETLSMEKIPPM VTTDQKQESKLEKDKASALQEKPLPEEKLIPEEEKIRSEKKPLLEKKPTPEDKLLPE AKTSAPEEQKHDLLKSQVQIAEEKLEGRVAPKTQVQEGKQPTKMEGLPSGTPQSLPKEDDKT KTIKEQPPPTAKPDQEKEDDKSDTSSSQPKSPQGLSDTGYSDDGISSSLGEIPLIPTDE KDILKGLKDSFSQESSPSSPSDLAKLESTVLSILEAQASTLADEKSEKTPQHEVSPQPKD QEKTSLSLSETLEITISEEIKESQEERKDTFFKDSQDDIPSSKDHKEKSEFVDDITTRREPYD SVEESSESNPVPQRKRTSVGSSSSDEYKQEDSQSGSEEDFIRKQIIEMSADEDASGSED DEFIRNQLKEISSSTESQKKEETKGKGITAGKHRRLTRKSSTSIDEDAGRRHSWHDEDEAF DESPELKYRETKSQESELVVTGGGLRRFKTIELNSTIADKYSAESSQKTSLYFDEPELE MESLTDSPEDRSRGGSSSLHASFPTGTSPTSVSLLDESDSSPSHKKGESKQQRKARHRPH GPLLPTIEDSSEEEELREEELKEQEQREIEQQQRKSSSKSKKDKDELRAQRRRRPKTP PSNLSPIEDASPTIELRQAAMEELHRSSCEYSPSIESDPEGFEISPEKIEVQVYKPLPTA VSLYSPTDEQSIMQKEGSQKALKSAEMEYEMMHKTHKYKAPPAANERDEVEFEKPLYGGM LI</p>

EDYIYESLVEDTYNGSVDSGLLTRQEEENGFMQQKGREQKIRLSEQIYEDPMQKITDLQKEFY ELESLSHVVQPEDIIVSSSFIIPESHEIVDLGTMVSTEEERKLLDADAAYEELMKRQQMQLTP GSSPTQAPIGEDMTESTMDFDRMPDASLTSSVLSGASLTDSTSSATLSIPDVKITQHFSTEEI EDEYVTDHTREIQEIIAHESLILTYSEPSESATSVPPSDTPSLTSSVSVCTTSSSSPITTLTD SITTVYTEPVDMITKFEDSEEISSSTYFPGSIIDYPEEISASLDRTPAPPDGRASADHIVISLS DMASSIIESVVPKPEGPVADTVSTDLLISEKDPVKKAKKETGNGIILEVLEAYRDKKELEAER TKSSLSETVFDHPPSSVIALPMKEQLSTTYFTSGETFGQEKPASQLPSGSPSVSSLPAPKPRPF FRSSSLDISAQPPPPPPPPPPPPPPPPPPPPPLPPPTSPKPTILPKKKLTVASPVTTATPLF DAVTTLETAVLRNGLPVTRICTTAPPPVPPKPSIPSGLVFTHRPEPSKPPAPKPVIPQL PTTQKPTDIHPKPTGLSLTSSMTLNSVTSADYKLPSPTSPLSPHSNKSSPRFSKSLTETYVV ITLPEPGTPTDSSASQAITSWPLGSPSKDLVSVFVSVVPPVTAVEIPISSEQTFYISGAL QTFSATPVTAPSSFOAAPTSTVQFLTTEVSKTEVSATRSTAPSVGLSSISITIPPEPLALDNI HLEKPYKEDGKLQLVGDVIDLRTVPKVEVKTDDKCIDLSASTMDVKRQITANEVYQKQISAV QPSIINLSVTSSIVTPVSLATETVTFVTCTASASYTTGTESLVGAEHAMTTPQLTTSKHAEP PYRIPSDQVFPPIAREEAPINLSLGTFAHAVTLAITKPVTVPPVGVNWTDSVSGITDGEV VDLSTTKSHRTVVTMDESTSSVMTKIIEDEKPVDLTAGRAVCCDVVYKLPFGRSCTAQQPAT TLPEDRFGYRDDHYQYDRSGPYGYRGIGGMKPSMSDTNLAEGHFFYKSKNAFYSEGDTAV DLTSGRVTTEGVMDYSSKTTGPYPETROVISGAGISTPQYSTARMTPPPGPQYCVGSVLRSSN GVVYSSVATPTPSTFAITTPQGSIFSTTVRDLSGIHTADAVTSLPAMHHSQMPRSYFITTTGA SETDIAVTGIDISASLQTTIMESLTAETIDSVPTLTASEVFEVVGDESALLIVPEEDKQQQ QLDLERELLELEKIKQORFAEELEWERQEIQRFREQEKIMVQKKLEELQSMQHLFLQEEER QAQFMMRQETLAQQQLQLEQIQQLQQQLHQLEEQKIRQIYQYNDPSGTASPTTTTEQAILE GQYAALEGSQFWATEDATTTASAVVAIEIPQSQGWYTVQSDGVTVQYIAPPGLISTVSEIPLTD VVVKEEKQPKKRSSGAKVRGQYDDMGENTDDPRSFKKIVDSGVQTDDEDATDRSYVSRRRRT KKSVDTSVQTDDEDQDEWDMPTSRRRKARVGKYGDSMTEADTKPLSKVSSIAVQTVAEISVQ TEPVGTIRTPSIRARVDAKVEIIKHISAPEKTYKGGSLGCQTEADSDTQSPQYLSATSPPKDK KRPTPLEIGYSSHLRADSTVQLAPSPPKSPKVLVSPISPLSPGKALESFVPEKPLPDDISP QKVLHPDMAKVPASPPTAKMMQSRMSDPKPLSPTADESSRAFFQYTEGYTTKGSQMTSSGA QKKVKRTLNPPEEISTGTQSTFSTMGTVSRRRICRTNTMARAKILQDIDRELDLVERESAK LRKKQAELEEEKEIDAKLRYLEMGINRRKEALLKEREKRERAYLQGVADRDYMSDSEVSST RPTRIESQHGIERPTAPQTEFSQFIPPTQTESQLVPPTSPYTQYQYSSPALPTQAPTSTYTQ QSHFEQQTLYHQVSPYQQTFTQAVATMSFTPQVQPTPTPQPSYQLPSQMMVVIQKPRQTTL YLEPKITSNYEVIRNQPLMIAPVSTDNTFAVSHLGSKYNSLDLRIGLEERSMSSPSSISA DSFYADIDHHTPRNYVLIDDIGEITKGTAALSTAFSLHEKDLSKTDRLRLTTETRRSQEVTFD LAPLQSSSRLHSYVKAEEDEMEDPYELKLLKHQIKQEFRRGTESLDHLAGLSHYHADTSYRH FPKSEKYSISRLTLEKQAAKQLPAAILYQKQSKHKKSLIDPKMSKFSPIQESRDLEPDYSSYM TSSTSSIGGISSRARLLQDDITFGLRKNI TDQKFMGSSSLGTGLTGLNTIRSAHQDEADKPY SSGSRSRPSSRPSSVYGLDLSIKRDSSSSSRLKAQEAELDVFSHASSSARTKPTSLPISQ SRGRIPIVAQNSEESPLSPVGQPMGMARAAAGPLPISADTRDQFGSSHSLEVEQQHMREES RTRGYDRDIAFIMDDFQHAMSDEAYHLRREETDWFDPKPRSRLENGHLDRKLPERLVHSRP LSQHQQEQIIQMNGKTMHYIFPHARIKITRDSKDHTVSGNGLGIRIVGGKEIPGHSGEIGAYIA KILPGGSAEQTGKLMEGMQVLEWNGIPLTSKTYEEVQSIISQSGEAEICVRLDLNMLSDSEN SQHLELHEPPKAVDKAKSPGVDPKQLAELQKVSLLQSSPLVLSVVEKGSHVHSGPTSAGSSS VPSPGQGPSVSKKKHGSKPTDGTGVVSHPTGEIQLQINYDLGNLIHILQARNLVPRDN NGYSDPFVKVYLLPGRGQVMVQNASAEYKRRTKHVQKSLNPEWNQTVIYKSI SMEQLKKTL EVTVWDYDRFSSNDFLGEVLIDLSSSTSHLDNTPRWYPLKEQTESIDHGKSHSSQSSQSPKPS VIKRSRSHGIFPDPSKMDQVPTIEKSHSSPGSSKSSSEGLRSHGSPRSQSSTSVTQTHLEDAG AAIAAAEAAVQQLRIQPSKRRK

Further analysis of the NOV65a protein yielded the following properties shown in Table 65B.

Table 65B. Protein Sequence Properties NOV65a	
PSort analysis:	0.9800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV65a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 65C.

Table 65C. Geneseq Results for NOV65a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV65a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM58340	Human brain expressed single exon probe encoded protein SEQ ID NO: 30445 - Homo sapiens, 748 aa. [WO200157275-A2, 09-AUG-2001]	1278..2025 1..748	748/748 (100%) 748/748 (100%)	0.0
AAM58328	Human brain expressed single exon probe encoded protein SEQ ID NO: 30433 - Homo sapiens, 396 aa. [WO200157275-A2, 09-AUG-2001]	3976..4371 1..396	396/396 (100%) 396/396 (100%)	0.0
AAM58208	Human brain expressed single exon probe encoded protein SEQ ID NO: 30313 - Homo sapiens, 182 aa. [WO200157275-A2, 09-AUG-2001]	2789..2970 1..182	182/182 (100%) 182/182 (100%)	e-103
AAM32174	Peptide #6211 encoded by probe for measuring placental gene expression - Homo sapiens, 162 aa. [WO200157272-A2, 09-AUG-2001]	1823..1984 1..162	162/162 (100%) 162/162 (100%)	2e-87
AAM71894	Human bone marrow expressed probe encoded protein SEQ ID NO: 32200 - Homo sapiens, 162 aa. [WO200157276-A2, 09-AUG-2001]	1823..1984 1..162	162/162 (100%) 162/162 (100%)	2e-87

5

In a BLAST search of public sequence databases, the NOV65a protein was found to have homology to the proteins shown in the BLASTP data in Table 65D.

Table 65D. Public BLASTP Results for NOV65a				
Protein Accession Number	Protein/Organism/Length	NOV65a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y6V0	Piccolo protein (Aczonin) (DJ0897G10.1) - Homo sapiens (Human), 5147 aa (fragment).	37..4872 1..4760	4753/4836 (98%) 4756/4836 (98%)	0.0

Q9JKS6	Piccolo protein (Multidomain presynaptic cytomatrix protein) - <i>Rattus norvegicus</i> (Rat), 5085 aa.	1..4870 1..4877	4060/4924 (82%) 4317/4924 (87%)	0.0
Q9QYX7	Piccolo protein (Presynaptic cytomatrix protein) (Aczonin) (Brain- derived HLMN protein) - <i>Mus musculus</i> (Mouse), 5038 aa.	1..4870 1..4830	4047/4922 (82%) 4293/4922 (86%)	0.0
Q9PU36	Piccolo protein (Aczonin) - <i>Gallus gallus</i> (Chicken), 5120 aa (fragment).	87..4872 1..4853	3260/4986 (65%) 3724/4986 (74%)	0.0
T00332	hypothetical protein KIAA0559 - human, 1212 aa (fragment).	3662..4873 1..1212	1212/1212 (100%) 1212/1212 (100%)	0.0

Pfam analysis predicts that the NOV65a protein contains the domains shown in the Table 65E.

Table 65E. Domain Analysis of NOV65a			
Pfam Domain	NOV65a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PDZ	4440..4527	26/90 (29%) 69/90 (77%)	8.4e-09
C2	4647..4745	40/104 (38%) 78/104 (75%)	1.9e-27

5

Example 66.

The NOV66 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 66A.

Table 66A. NOV66 Sequence Analysis			
	SEQ ID NO: 241	2914 bp	
NOV66a, CG138208-01 DNA Sequence	CCGGGCCATGGCGGAACGCGAGGGGCGGGCGGTGGTCCCGGAGGCGCCGGGGCGGCAGCGG CCAGCGGGGATCCGGGGTCGCCCAGTCCCTCAGCAGCCGCCCGCAGCAGCAGCAGCAGCA GCCGCCGAGCAGCCGACGCCCCCAAGCTGGCCAGGCCACCTCGTCGTCTCTGTCACCTC GGCGGGCGGTGCTCCTCCTCGTCTCGTCTACCTCCACCTCCATGGCCGTGGCGGTGGCCCTC GGGCTCCGCGCCTCCCGGTGGCCCGGGCCAGGCCGACCCCCCGCCCGGTGCAGATGAACCT GTACGCCACCTGGGAGGTGGACCGGAGCTCGTCCAGCTGCGTGCCCTAGGCTATTGAGCTTGAC CCTGAAGAACTCGTCATGCTAAAAGAAATGGACAAAGATCTTAAGTCAAGTGGTTCATCGCTGT GAAGCTGAGGGTTCAAAAAGAAATCTTCGCTCCAACGAGATCGTCTTCCAGCTAGTGGACT GGTGGAAACAGAGCTCCAATTAACCTTCTCCCTTCAGTACCCTCATTTCCTTAAGCGAGATGC CAACAAGCTGCAGATCATGTGCAAGGAGAGAAAACGTTACAAGAATCGGACCATCTTGGGCTA TAAGACCTTGGCCGTGGGACTCATCAACATGGCAGAGGTGATGCAGCATCCTAATGAAGGCGC ACTGGTCTTGGCTACACAGCAACGTGAAGGATGTCTCTGTGCTGTGGCAGAAATAAGAT CTACTCCTGTCCAGCCAACCATTTGACCATGAAGGAATCAAATCCAAGCTTTCTGATCGTTT TCCTGATATTGACAATTATTCTGAGGAAGAGAGAGTTTCTCATCAGAACAGGAAGGCAG TGATGATCCATTGCATGGGCAGGACTTGTCTACGAAGACGAAGATCTCCGGAAGTGAAGAA GACCCGGAGGAACTAACCTCAACCTCTGCCATCACAAGCAACCTAACATCAAACAGAAGTT		

	<p>TGTGGCCCTCCTGAAGCGTTTAAAGTTTCAGATGAGGTGGGCTTTGGGCTGGAGCATGTGTC CCGCGAGCAGATCCGGGAAGTGAAGAGGACTTGGATGAATTGTATGACAGTCTGGAGATGTA CAACCCAGCGACAGTGGCCCTGAGATGGAGGAGACAGAAAGCATCCTCAGCAGCCAAAGCC CAAGCTCAAGCCTTTCTTTGAGGGGATGTGCGAGTCCAGCTCCAGACGGAGATTGGCAGCCT CAACAGCAAAGGCAGCCTCGGAAAAGACACCAGCCCTATGGAATTGGCTGCTCTAGAAAA AATTAAATCTACTTTGGATTAAAAACCAAGATGACAGCTTGACTGAAAACAGACACTCTGGAAT CACTGACCAGGACATGTTTGGAGATGCCAGCAGAGTCTGGTGTGCCGGAGAAAGTCAAAAC TCCATGAAGTCCAGTAAACGGATCTCCAGGGCTCTGCCTCCCCAGCAAAGTGGAGGGGGT GCACACACCCCGGCAGAAGAGGAGCAGCCCTGAAGGAGCGGCAGCTCTCAAAGCCCTTAAG TGAGAGGACCAACAGTTCGACAGCGAGCGTCCCGAGATCTGGGCCACAGCAGCAGATTCC AAGAAAGGTGGTGTATGACCAGCTCAATCAGATCCTGGTGTGATGCAGCCCTCCAGAAAA TGTCACTTCTGGTGAACACCAGTACTGGCAGGGCCAGTATGTGGCTGAGCTGCTCCAGGACCA GCGGAAGCCTGTGGTGTGCACCTGCTCCACCGTGGAGGTCCAGGCCGTGCTGCTCCGCCCTGCT CACCCTGGATCCAGCGCTACTGCAACTGCAACTCTTCCATGCCAGGCCAGTGAAGGTGGCTGC TGTGGGAGGCCAGAGCTACCTGAGCTCCATCCTCAGGTCTTTGTCAAGTCCCTGGCCAAAC GACCTCCGACTGGCTTGGCTACATGCGCTTCTCATCATCCCCCTCGGTTCTCACCCTGTGGC CAAATACTTGGGGTCACTGACAGTAAATACAGTAGTTCCTTCTGGATTCTGGTTGGAGAGA TCTGTTCACTGCTCGGAGCCACAGTGTGACAGCAACTGGACGTGGCAGGGCGGGTGTATGCA GTACGTCAACGGGGCAGCCACGACACACCAGCTTCCGTGGCCGAAGCATGCTGACTTGCCG GCATAAGTTCCCTGATGAAGACTCCTATCAGAAAGTTATTCCCTTCATTGGCGTGGTGAAGT GGGTCTGGTTGAAGACTCTCCCTCCACAGCAGGCGATGGGGACGATTCTCTGTGGTCAGCCT TACTGTGCCCTCCACATCACCACCTCCAGCTCGGGCTGAGCCGAGACGCCAGGCCACCCCC TCCCTCCTCCCCATCTATGAGCAGCGCCCTGGCCATCGTGGGGAGCCCTAATAGCCCATATGG GGACGTGATTGGCCCTCCAGGTGGACTACTGGCTGGGCCACCCCGGGGAGCGGAGGGAAGG CGACAAGAGGGACGCCAGCTCGAAGAACACCCTCAAGAGTGTCTTCCGCTCAGTGCAGGTGTC CCGCTTCCCCATAGTGGGGAGGCCAGCTTCTGGCACCATGGCCATGACTGTGGTCACCAA AGAAAAGAACAAGAAAGTCTCCACCATCTTCTGAGCAAGAAACCCGAGAAAGGAGTGTGA TTCTAAGAGCCAGGTCAATGAAGGCATCAGCGCCTCATCTGCTCAGCAAGCAGCAGAGAC TATGCTGAGAGTGTCCATCGATGGGGTCGAGTGGAGTGACATCAAGTCTTCCAGCTGGCAGC CCAGTGGCCACCCATGTCAAGCACTTCCAGTGGGACTCTTCACTGGCAGCAAGGCCACCTG AGGCCCTGTCTCCAG</p>
	<p>ORF Start: ATG at 8</p> <p>ORF Stop: TGA at 2897</p>
	<p>SEQ ID NO: 242</p> <p>963 aa</p> <p>MW at 104897.3kD</p>
NOV66a, CG138208-01 Protein Sequence	<p>MAERGGAGGGPGGAGGGSGQRGSGVAQSPQQPPPPQQQQQQPPQPTPPKLAQATSSSSSTSA AASSSSSSTSTSMVAVASGSAPPGGPGPRTAPVQMNLYATWEVDRSSSSCVRLFSLTLK KLVLKEMDKDLNSVVIIVKLGSKRILRSNEIVLPASGLVETELQLTFLQYPHFLKRDANK LQIMLQRRKRYKNRTILGYKTLAVGLINMAEVMQHPNEGALVGLHSNVKDVSPVAEIKIYS LSSQPIDHEGIKSLSDRSPDIDNYSSEEEESFSSEEGSDPLHGQDLFYEDDLRKVKKTR RKLTSTSAITRQPNIKQKFVALLKRFKVSDEVGFGLEHVSREQIREVEEDLDELYDSLEMYNP SDSGPEMEETESILSTPKPKLKPFEGMSQSSSQTEIGSLNSKSGSLGKDTTSPMELAALEKIK STWIKNQDDSLTETDLEITDQDMFGDASTSLVPEKVKTPMKSSKTDLQGSASPSKVEGVHT PRQKRSTPLKERQLSKPLDALERTNNSDSERSPDLGHSSTQIPRKVVYDQLNQILVSDAALPENV LVNTTDWQGYVAELLQDQKPVVCTCSTVEVQAVLSALLTRIQRVCNCSMPRPVKAAGV GQSYLSSILRFFVKSLANKTSDWLGYMFLIPLGSHPVAKYLGVSVDKYSFSLDSGWRDLF SRSEPPVSEQLDVAGRVQMVGNAATTHQLPVAEAMLTCHKFPDEDSYQKFIPIFVGVKVL VEDSPSTAGDGDSPVVSILTVPSTSPSSSGLSRDATATPPSSPSMSALAIUGSPNSPYGDL IGLQVDYWLGHGERRREGDKRDASSKNTLKSIVRSVQVSRPLPHSGEALSGTMAMTVVTKEK NKKVPTIFLSKKPREKEVDSKSQVIEGIRLICSQKQQTMLRVSIDGVEWSDIKFFQLAAQW PTHVKHFPVGLFSGSKAT</p>
	<p>SEQ ID NO: 243</p> <p>3225 bp</p>
NOV66b, CG138208-02 DNA Sequence	<p>CGGCCTCCGTAACCCCCGCTAGCCGGGCGCATGGCGGAACCGGAGGGGCGGGCGGTGCTCCC GGAGGCGCGGGGGCGGCAGCGGCCAGCGGGGATCCGGGGTGCAGCAGTCCCTCAGCAGCCG CCGCGCGCAGCAGCAGCAGCAGCAGCGCGCGCAGCAGCCAGCGCCCCAAGCTGGCCAGCC ACCTCGTCTCCTCGTCCACCTCGGCGGGCGGTGCTCCTCCTCGTCTCCTCGTCTCCTCACC TCCATGGCCGTGGCGGTGGCCTCGGGCTCCGCGCCTCCCGGTGGCCCGGGGCCAGGCCGCACC CCGCCCCCGGTGCAGATGAACCTGTACGCCACCTGGGAGGTGGACCGGAGCTCGTCCAGCTGC GTGCCTAGGCTATTACGCTTGACCTGAAGAACTCGTCTATGCTAAAAGAAATGGACAAAGAT CTAACTCAGTGGTCATCGTGTGAAGCTGCAGGGTTCAAAAAGAAATTCCTCGCTCCACAGAG ATCGTCTCTCCAGCTAGTGGACTGGTGGAAACAGAGCTCCAATTAACCTTCTCCCTCAGTAC CCTCATTCTCTTAAGCGAGATGCCAACAGCTGCAGATCATGCTGCAAAGGAGAAAACGTTAC AAGAATCGGACCATCTTGGGCTATAAGACCTTGGCCGTGGGACTCATCAACATGGCAGAGGTG ATGCAGCATCCTAATGAAGGCGCACTGGTGCTTGGCCTACACAGCAACGTGAAGGATGTCTCT GTGCCCTGGGCAGAAATAAAGATCTACTCCTGTCCAGCCAACCCATTGACCATGAAGGAATC AAATCCAAGCTTTCTGATCGTTCTCTGATATTGACAATTATTCTGAGGAAGAGGAAGAGAGT</p>

	<p>TTCTCATCAGAACAGGAAGGCAGTGTATGCCATTGCATGGGCAGGACTTGTCTACGAAGAC GAAGATCTCCGGAAGTGAAGAAGACCCGAGGAACTAACCTCAACCTCTGCCATCACAAAGG CAACCTAACATCAACAGAAAGTTTGTGGCCCTCCTGAAGCGGTTAAAGTTTCAGATGAGGTG GGCTTTGGGCTGGAGCATGTGTCCCGCAGCAGATCCGGGAAGTGAAGAGGACTTGGATGAA TTGTATGACAGTCTGGAGATGTACAACCCAGCGACAGTGGCCCTGAGATGGAGGAGACAGAA AGCATCCTCAGCACGCCAAAGCCCAAGCTCAAGCCTTCTTTGAGGGGATGTCCAGTCCAGC TCCCAGACGGAGATTGGCAGCCTCAACAGCAAAGGCAGCCTCGGAAAAGACACCAGCCCT ATGGAATTGGCTGTCTAGAAAAAATAAATCTACTTGGATTAAAAACCAAGATGACAGCTTG ACTGAAACAGACACTCTGGAATCACTGACCAGGACATGTTTGGAGATGCCAGCAGAGTCTG GTTGTGCCGAGAAAGTCAAACTCCCATGAAGTCCAGTAAACCGGATCTCCAGGGCTCTGCC TCCCCAGCAAAGTGGAGGGGGTGACACACCCCGGAGAAAGAGGAGCAGCCCTGAAGGAG CGGCAGCTCTCAAGCCCTAAGTGAAGGACCAACAGTTCGACAGCGAGCGTCCCCAGAT CTGGGCCACAGCAGCAGATTCCAAGAAAGGTGGTGTATGACCAGCTCAATCAGATCCTGGTG TCAGATGCAGCCCTCCAGAAAATGTCATTCTGGTGAACACCACTGACTGGCAGGGCCAGTAT GTGGCTGAGCTGCTCCAGGACCGGGAAGCCTGTGGTGTGCACCTGTCCACCGTGAAGGTC CAGGCCGTGTGTCCGCCCTGTCTACCCGGATCCAGCGCTACTGCAACTGCAACTCTTCCATG CCGAGGCCAGTGAAGGTGGTGTGTGGGAGGCCAGAGCTACCTGAGTCCATCCTCAGGTTT TTTGTCAAGTCCCTGGCCAAACAAGACCTCCGACTGGCTTGGCTACATGCGCTTCTCATCATC CCCTCGGTTCTCACCTGTGGCCAAATACTTGGGGTCAGTCGACAGTAAATACAGTAGTTCC TTCTTGGATTCTGGTTGGAGAGATCTGTTCACTCGCTCGGAGCCACCACTGTGAGCAACTG GACGTGGCAGGGCGGGTGTGTCAGTACGTCAACGGGGCAGCCAGACACACCAGCTTCCCGTG GCCGAAGCCATGTGACTTGCCGGCATAAGTTCCTGATGAAGACTCCTATCAGAAGTTATT CCCTTCATTGGCGTGGTGAAGGTGGTCTGGTGAAGACTCTCCCTCCACAGCAGGGCAGTGGG GACGATTCTCCTGTGGTCAGCCTTACTGTGCCCTCCACATCACACCTCCAGCTCGGGCTG AGCCGAGACGCCACGGCCACCCCTCCCTCCTCCCATCTATGAGCAGCGCCCTGGCCATCGTG GGGAGCCCTAATAGCCCATATGGGGACGTGATTGGCTCCAGGTGGACTACTGGCTGGGCCAC CCCGGGGAGCGGAGGAGGGAAGGCGACAAGAGGGACGCCAGCTCGAAGAACACCCCTCAAGAGT GTCTTCCGCTCAGTGCAGGTGTCCCGCCTGCCCATAGTGGGGAGGCCAGCTTCTGGCCAC ATGGCCATGACTGTGGTCACCAAGAAAAGAAACAAGAAAGTTCACCATCTTCTGAGCAAG AAACCCCGAGAAAAGGAGGTGGATTCTAAGAGCCAGGTCAATGAAGGCATCAGCCGCTCATC TGTTCTTCCCCCTCCTTAGGCCCCAGCCTGGGCCCAGACCCATCCTCCAGCCAGGTTTCCCT CCAGCAGGCTCCTTCCCTCCCTGTACCTCCCTCTACCAACCCGGGTCTGAGCCCTCATT CCTGACCGTCCGTGTTCTCAGAGTGGTTGAGGACACAGGGCCCCAGCCAGCCCTCTGCACC CCCCAGCCCGCCATCTGCGCCCCACAGCCCTTTGGAGCTTTTCTTGTCTCTCACTCCT TCCAGAAAGTTTTTGACAGAACTTCATTTGAAAGTGTTTTTCTCATTCTCCATACCTCCCC CAAGCTCTCCTCCAGCCCTTCCAGGGCTCAGCCCTGCTGTCTGAGCGTCTCCTGGGCCAGA GAGAGGATGGGGGTGGGAGGACTGAGTTGATGTGGGTTTTTCTCATTCAATAAATTGGTGA TTTCTTACCGAC</p>
	<p>ORF Start: ATG at 31</p>
	<p>ORF Stop: TGA at 3055</p>
	<p>SEQ ID NO: 244</p>
	<p>1008 aa</p>
	<p>MW at 109341.2kD</p>
<p>NOV66b, CG138208-02 Protein Sequence</p>	<p>MAERGGAGGPGGAGGSGQSGVAQSPQPPPPQQQQQPPQOPTPPKLAQATSSSSSTSA AASSSSSTSTMAVAVASGSAPPGGPGGRTAPVQMNLYATWEVDRSSSSCVPRFLSLTLK KLVLKEMDKDLNSVVIIVKQSGSKRILRSNEIVLPASGLVETELQTLFSLQYPHFLKRDANK LQIMLQRRKRYKNRTILGYKTLAVGLINMAEVMQHPNEGALVLGLHSNVKDVSVPAEIKIYS LSSQPIDHEGIKSLSDRSPDIDNYSEEEESFSSEQEGSDPLHGQDLFYEDDLRKVKKTR RKLTSTSAITRQPNIKQKFVALLKRFKVSDEVGFLEHVSREQIREVEEDLDELDSLEMYNP SDSGPEMEETESILSTPKPKLPPFEGMSQSSSQTEIGSLNSKSLGKDTTSPMELAALEKIK STWIKNQDDSLTETDLEITDQDMFGDASTLVVPEKVKTPMKSSKTDLQGSASPSKVEGVHT PRQKRSTPLKERQLSKPLSERTNSSDSERSPD LGHSTQIPRKVVYDQLNQLVSDAALPENVI LVNTTDWQGYVAELLQDQRPVCTCSTVEVQAVLSALLTRIQRVCNCSMPRPVKVAAG GQSYLSSILRFFVKSLANKTSDWLGMYRFLIPLGSHPVAKYLGVSVDISKYSSFLDSGWRDLF SRSEPPVSEQLDVAGRMQYVNGAATTHQLPVAEAMLTCHKFPDEDSYQKFIPIFIVVKVGL VEDSPSTAGDGDSPVSLTVPSTSPSSSGLSRDATATPPSSPSMSALAVGSPNSPYGDV IGLQVDYWLGHHPGERRREGDKRDASSKNTLKSIVRSVQVSRPLPHSGEQLSGTMAMTVVTKK NKKVPTIFLSKKPREKEVDSKSQVIEGISRLICSSPSLGPSPGDPSSQPGFPAGSPPCHL PLTNPGSEPLIPDRPCSQEWLRTQGPSPALCTPQPGHLRPTAPLELFSCLPTPSQKFLHRTSF</p>

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 66B.

Table 66B. Comparison of NOV66a against NOV66b.		
Protein Sequence	NOV66a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV66b	100..915	738/816 (90%)
	100..915	738/816 (90%)

Further analysis of the NOV66a protein yielded the following properties shown in Table 66C.

Table 66C. Protein Sequence Properties NOV66a	
PSort analysis:	0.9700 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV66a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 66D.

Table 66D. Geneseq Results for NOV66a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV66a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB42247	Human ORFX ORF2011 polypeptide sequence SEQ ID NO:4022 - Homo sapiens, 885 aa. [WO200058473-A2, 05-OCT-2000]	85..963 7..885	879/879 (100%) 879/879 (100%)	0.0
AAB92606	Human protein sequence SEQ ID NO:10867 - Homo sapiens, 389 aa. [EPI074617-A2, 07-FEB-2001]	620..917 1..298	296/298 (99%) 297/298 (99%)	e-168
AAB57122	Human prostate cancer antigen protein sequence SEQ ID NO:1700 - Homo sapiens, 543 aa. [WO200055174-A1, 21-SEP-2000]	468..963 48..542	289/501 (57%) 371/501 (73%)	e-159
ABB61495	Drosophila melanogaster polypeptide SEQ ID NO 11277 - Drosophila melanogaster, 1164 aa. [WO200171042-A2, 27-SEP-2001]	59..956 34..1152	336/1140 (29%) 490/1140 (42%)	e-101
AAB38002				8e-09

	gene 19 clone HCUGE72 - Homo sapiens, 104 aa. [WO200055371-A1, 21-SEP-2000]	65..102	32/38 (83%)	
--	---	---------	-------------	--

In a BLAST search of public sequence databases, the NOV66a protein was found to have homology to the proteins shown in the BLASTP data in Table 66E.

Table 66E. Public BLASTP Results for NOV66a				
Protein Accession Number	Protein/Organism/Length	NOV66a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O88588	Cytosolic sorting protein PACS-1a - Rattus norvegicus (Rat), 961 aa.	1..963 1..961	940/963 (97%) 948/963 (97%)	0.0
BAC04831	CDNA FLJ39325 fis, clone OCBBF2015334, highly similar to Rattus norvegicus cytosolic sorting protein PACS-1a (PACS-1) mRNA - Homo sapiens (Human), 829 aa (fragment).	1..879 1..829	825/879 (93%) 826/879 (93%)	0.0
Q96MW0	CDNA FLJ31799 fis, clone NT2R12009037, highly similar to Rattus norvegicus cytosolic sorting protein PACS-1a (PACS-1) mRNA - Homo sapiens (Human), 750 aa (fragment).	130..879 1..750	750/750 (100%) 750/750 (100%)	0.0
Q9ULP5	KIAA1175 protein - Homo sapiens (Human), 641 aa (fragment).	323..963 1..641	641/641 (100%) 641/641 (100%)	0.0
O88589	Cytosolic sorting protein PACS-1b - Rattus norvegicus (Rat), 559 aa.	1..543 1..541	532/543 (97%) 535/543 (97%)	0.0

5

PFam analysis predicts that the NOV66a protein contains the domains shown in the Table 66F.

Table 66F. Domain Analysis of NOV66a			
Pfam Domain	NOV66a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 67.

The NOV67 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 67A.

Table 67A. NOV67 Sequence Analysis			
	SEQ ID NO: 245	5634 bp	
NOV67a. CG138303-01 DNA Sequence	CGGGATTGCACCATGGGGAACCAAGGATGGGAAGCTGAAGAGGAGCGCAGGTGATGCTTTGCAC GAAGGCGGCGGTGGCGCCGAGGATGCGCTGGGGCCCAGGGATGTGGAAGCCACAAAGAAGGGG AGCGGGGGCAAGAAGGCGCTAGGCAAGCACGGCAAGGGGGAGGGGGCGGCGGCGGCGCGG GAGTCGGGCAAGAAGAAGAGCAAGTCCGACTCCAGAGCCTCGGTGTTTTCAACCTGCGGATC AGGAAGAATCTGTCCAAGGGGAAAGGCGCGGCGGCTCCCGCAAGATGTACTGGATTCCAG GCCCTGCAGACCGGGGAGCTGGACAGCGCTCACTCCCTGCTCACCAAGACTCCAGACCTCAGC CTCTCGGCGGACGAGCCGGCTGTGCGATACCGAGTGTGCGGACCCTTTTGAGGTGACCGGT CCAGGGGGTCTTGGGCTGCCGAGGCTAGGGTCGGGGGCGGGCGATCGCCGAGGATGTGGAA ACTGCAGCAGGGGCGCAGGATGGACAAAGGACCAGCTCGGGCTCGGACACGGACATCTATAGC TTCCATTTCGGCTACGGAGCAAGAGGATTTGCTTTCAGACATCCAGCAGCGATCCGCTGCAG CAGCAGCAGCAGCAGCAGCTCCAGCTCCAGCTCCAGCAACAGCAGCAGCAGCAGCAGCTCCAG GGCGCCGAGGAGCCTGCAGCGCCCCCACTGCGGTCTCCCTCAGCCCCGGGCGCTTCTGGGC CTGGACCGGTTCTGTGCTGGGGCCGGTCTCCGAGGCGCCAGTCTCCGGCAGCGCAACCCGCG GCCAAAGACTCGCCCTCCTCCACGGCTTTCCTATTTCCGAGGCGGGGCGGGGAGGAACG GCCGGAGCCCCGTGCGAGGGGCTGGGGACACGGATGAGGAGGGTGAGGAGGATGCTTTTGAG GATGCGCCCCGGGGCTCTCCGGGGGAGGAGTGGGCCCGGAGGTGGGAGAGGACGCCCCGAG AGGCTGGGGGAAGAGCCGGAGGAGGAGGCGCAAGGACCTGACGCCCCGCGGCGCTTCCCTG CCGGCAGCCCCGCGCTAGCCAGCGCTGTTTCAAGCCCTACCCGCTCATCACCCCTGCTAC ATCAAGACCACCACCGGAGCTCAGCTCGCCAATCACTCCCGTCTCAGTCCCTTAATCAG AGCCCCAGGATCAAGAGGCGGCGGAACCTCCCTGAGCCGAGGGTCCAGAATGCCCTGGCC TCCGTAGCCGCCCCGCGCAAGAAGCAGCGGGCGGACGCGGCGCTTGCGGCGGCGCTGAGCCGC TCGGCTGACTGGACGAGGAGCTAGGCGCCCGCACGCCCCGGGTGGGAGGCTCCGCGCACCTG CTGGAGCGCGGGGTGGCGAGTGACAGCGCGGTGGGTGTCCCAGCAGCTGGCCGCGCAAGCG TCTGGGGCCCCCGGCTGCGGATGGCTTCCAGAACGTGTTACAGGTGAGCGCGGGCGAAGC CTGTTGGAGAAGCTGTTAGCCAGCAGGAGAACGGGCTCCAGAAGAAGCAGAGAAGTTTTC TCCCGGATCATTGCCATGGGTCTTCTCCTTCTTTAGTGATTGCTTCAGGGAACCGTGTAA CAGAATGCCCAAGATCAACTTTATACCTGGGCTGCAGTTAGTCAACCCACACTCATTGGAC TATTAGAAAGGGCAGTTTCTTAGGCGAGTTCCATCCATGGGGCCACCATCCAAACCTCCCGAT GAGGAACACAGGCTCGAGGATGCTGAAACAGAATCTCAATCTGCTGTTTCAGAACTCCCCAA AAACGCTCAGATGCTGTCCAGCAGGAAGTTGTTGACATGAAGTCTGAGGGACAGGCCACTGTA ATTACAGCAGCTGGAACAGACTATTGAGGATCTGAGAACAAAATAGCTGAAGTAGAGAGGCAG TATCTGCCCTGGACAGAGAGTGGCGAGTGGTTCATCAAGGGCTTGAGAATGGAGTACAGCC TCAGGCGATGTCTGTCTGAAGCTCTCAGGTTAGAAGAAAAGGAAGTACGGCATCATAGGATT TTAGAGGCGAAATCGATACAGACTTCCCCACGGAAGAGGGCGGGGTGTGACACTGCCTCCT GTGGATGGGCTGCCAGGGCGTCTCCATGCCCCCTGGGGCTGAAAGTGGACCTCAGACAAAG TTCTGTTTCAAGATTTCTTTGATTGTGTCTCAAGGCGAATATCAGTCCAGCTCGACAGCCAT CAGCCACACAGAGCATCTCAGAGCTCCACCACCTCCATCCCTTCTGTGTCTGCTGGGCAA GGACAGCTGGGTACAGCGCCCCATTCTATTTCTACCGAGTTTCAAACAGCCACGAACAC TCTGTTTCTCTGCTTTAAAAACAGCTGTAACATCCCATCTCCACCACCTCTGCCTTGACACA GAGTCTCCAGCTCCATGCTTGGCTGGGCTGGGATGGTGCTCCCCACCTCCCCCTCTCCCTGGC ATGACAGTGCCTACTCTGCCAGTACAGCCATTCCCCAACCTCTCTCTCAGAGGTACAGAA ATGCTGCCACCCCTCCCCCTCTTCTCCGGAGCGGGCATACTCTCCGCGGCTCTACCC GGAGCAGGCATCTCCCTCTGGGAGCGGGCATACCCCACCTCCCCCTCTACCGGAGCGGGC ATACCCCTCCGCCCCCTTACCCGGAGTGGGCATACCTCTCCGCCCCCTTACCCGGAGCG GGCATAACCCCTCTCCCCCTTACCCGGAGCGGGCATAACCCCTCTCCCCCTCTTCCCGGA GCGGGCATACTCTCCACCCCTTACCCAGAGTGGGCATAACCCCTCCGCCCCCTTCCCTCC GGAGCGGCATAACCCACCTCCCCCTTACCCGGAGCGGGCATAACCCCTCCGCCCCCTCTA CCTGGAGTGGGAATACCTCTCCGCCCCCTTACCTGGAGTGGGAATACCTCTCCGCCCCCT CTACCTGGTGTGGGATCCCCACCTCTCCCTTGGCAGGTATGGGGATTCCACCTGCTCCA GCTCCCCCACTCTCCCTGAGTGGGACAGGAATCCACCGCCCCCTCTGCTTCTGTATCAGGC CCTCCACTCTCCCAAGTTGGGAGTAGCACTTTACCAACCCACAGGTGTGTGGATTCTT CCTCCTCATTGCCAAGTGGCTTGTGGATTAGGGATGAATCAGGACAAAGGGAGTAGGAAG CAGCCCATAGAGCCTTGTGACCAATGAAGCCTCTTACTGGACAGGATCAACTACATAGT AAAAGAGACTCCAGTACTTCACTTATTTGGGAAAAAATTGAAGAGCCATCCATAGATTGTCAT GAATTTGAGGAATTATTTTCTAAAACCTGCTGTAAAGGAGAGAAAGAAACCTATCTCTGATACT		

	<p>ATCTCAAAGACGAAGGCTAAACAACAGGTTGTCAAGTTATTAAGCAACAAAAGATCACAAGCA GTTGGAATACTAATGTCTAGCCTTCATTTAGATATGAAAGACATACAACATGCTGTTGTGAAC TTGGATAATTCTGTGGTTGACCTGGAGACCCTTCAAGCTCTCTATGAGAATAGAGCACAGTCA GACGAACTCGAAAAATAGAAAAGCATGGCCGATCTTCCAAAGACAAGGAAATGCCAAGTCT CTGGACAAACCTGAACAGTTCCTTTATGAAGTGTCACTAATCCCAACTTTTCAGAGCGAGTCT TTTTGCATCCTGTTCCAGTCCACATTTTCAGAAAGCATTTGCTCAATTCGTCGCAAACTGGAA TTACTACAGAAATTTGTACATTAATAAATGGCCAGGGGTTATGCAGGTTCTAGGTTTGGTT CTTGCCCTTTGGCACTACATGAATGGAGGAAATAAGACTCGAGGACAGGCAGATGGCTTTGGA TTAGACATTTCTCCAAACTGAAAGATGTCAAGAGCAGTGACAATAGCAGAAGCCTTTTGTCA TATATTGTTTCGTATTATCTCCGAAATTTTGATGAGGATGCTGGAAAAGAACAGTGCCTCTTT CCACTGCCAGAACCCAGGACCTTTTTCAGGCCTCACAGATGAAGTTTGAAGATTTTCAAAAA GATCTCAGAAAACCTGAAGAAAGACTTGAAAGCCTGTGAAGTTGAAGCAGGGAAGTATACCAG GTCTCCTCAAAGAGCATATGCAGCCTTTTCAGGAAAACATGGAACAATTTATTATTCAAGCC AAAATTGACCAAGAGGCAGAGGAAATTCAGTACAGAGACTCATAAATGCTTTTGGAGACC ACGCATATTTCTTCATGAACCAAACTTGAGAGAGAAGGAGGTGTCCTCCAAATGCTTTCTTC AGTATCTGGCATGAATTCAGCTCTGACTTTAAAGACTTCTGGAAGAAAGAGAACAACTTCTT CTACAAGAGAGAGTAAAGAAAGCCGAGAGGTGTGTAGACAGAAAGAAAGGAAATCACTTTAT AAAATAAAACCAAGACATGACTCTGGAATTAAGCAAGATAAGCATGAAACTTGAACAATG AAAAGCAGAATGAAATGAGTCATTGCAACGACTTTTCAAAAATTTCAGCTGACCTGAGAGTGG GAGGGAACCTACCGTCATTCTGCTCATGTTTCTTCTTGACCTCTGCATAATCTTTTGT CTAGACAGTTCACAAATTTGTAATTTTACTGTATATTCATATAAAATGCAACGTAAGTACA CCAGTGGAGAATTTGACACCTTTTCTTTTGTAAAGTTTATGGTATTATACCGATAGACCAA AACAGCATGTGTAAGAGGCAGTATCTGCACATAATTCACATGCTAAACATTAACATAAT CACTGTTGTGAGAATATTCTCTGTCACAGCAAAACACTTTCTTTCTACTGACAAACAGTCC TCCACATCACAGCATTTAGACATATGGGTAAATGTTATTTCTAGTGAATTTGTTGTATCAGT TTCATGTCTAAGTATAAATTTCTATTTTAAATTTAAGAACCGTTTATAATCAGTGTCTTCC CAACTCTTGGGTTGCTCTCCATAACTATGTATTTGTGAAAGAAAATGGTCATTTTCTTACTG AAGTCATATAATGACTTGGGTGAGCTCGTAATGCATGTGTATGTTTGTATGAGCTGGGTGT TTTTTCCATTACTTTTAAATGATCTTCGTTGCAAGTTATAGTTGTGGATAAAGGGGAGAATTT ATTGCTCTTGCAACCAATATGGAAGCAACTTAAGAAAACCAATGTTCTAAATCATAATTG TTGTATTTATGTAAGATAGGTCTCTTACTTTTGTAGTTTGTAGTTTAAAGTCAAAGAAACAG TAGTGGTTTTTTTCTATTGTTTGTAGTCTTCTGTCCTTCAGTCCCTCCAGTGTGTATA TTACCATCTCCAATGAAATAATAGGGCATTTAACAAGATCGCTATGTGCAATCTGATTTT AGTGTCTTATTTCAATTTTCTAGGATGTTAATTTATATGAAATAAAATGAATAATAAAAG AATAAAGATAAAAAAAAAAAAAAAAAA</p>
	<p>ORF Start: ATG at 13</p>
	<p>ORF Stop: TGA at 4591</p>
	<p>SEQ ID NO: 246</p>
	<p>1526 aa</p>
	<p>MW at 162002.3kD</p>
<p>NOV67a, CG138303-01 Protein Sequence</p>	<p>MGNQDGKLRKRSAGDALHEGGGAEDALGPRDVEATKKGSGGKKALGKHGKGGGGGGGGESGK KKSKSDSRASVFNLRIRKNSLKGKAGGSREDVLDLQALQTGELDSAHSLLTPTDLSLSD EAGLSDETCADPFVETGPGGPGPAEARVGRPIAEDVETAGAQDQRTSSGSDTDIYFHSFA TEQEDLLSDIQAIRLQQQQQQQLQLLQQQQQQQLQGAEEPAAPPTAVSPQPGAFGLDLRF LLGPVSEAPSLPAAQPAADSPSTAFPFPEAGPGEAAAGAPVRGAGDTDEEGEEDAFEDAPR GSPGEEWAPEVGEDAPQRLGEEPEEEAQGPDAAPAAASLPSPAPSPQRCFKPYPLITPCYIKTT TRQLSSPNHSPSQSPNQSPRIKRRPEPSLSRGSRTALASVAAPAKKHRADGGLAAGLSRSADW TEELGARTPRVGSAAHLLERGVSDSGGVSPALAAKASGAPAAADGFQNVFTGERGRTLLEK LFSQQENGPPPEAEKFCRSRIAMGLLLPSDFREPCNQNAQDQLYTAAVVSQPTHSLDYSEG QFRRVPMSMGPPSKPPDEHRLEDAETESQSAVSETPQKRSDAVQQEVVDMKSEGQATVIQQL EQTIEDLRKIAELERQYPALDTEVASGHQGLENGVTASGDVLEALRLEEKEVRHHRILEAK SIQTSPTTEGGVLTLPVVDGLPGRPPCPGAESEGPQTKFCSEISLIVSPRRISVQLDSHQPTQ SISQPPPPPSLLWSAGQGQPGSPHSTEFQTSHEHSVSSAFKNSCNIPSPPLPCTESS SMPGLGMVPPPPPLPGMTVPTLPSTAIQPPPLQGTEMLPPPPPLPGAGIPPPPLPGAGI LPLGAGIPPPPLPGAGIPPPPLPGVGIPPPPPLPGAGIPPPPLPGAGIPPPPLPGAGIP PPPPPLRVGIPPPPLPGAGIPPPPLPGAGIPPPPLPGVGIPPPPPLPGVGIPPPPPLPGA GIPPPPLPGMGIPPAAPAPPLPPPGTGIPPPPLPVSGPPLLPQVGSSTLPTPQVCGFLPPLP PSGLFGLGMNQDKGSRKQPIEPCRPMPKPLYWTRIQLSKRDSSSTLIWEKIEPSIDCHEFEE LFSKTAVKERRKPISDTISKTKAKQVVKLLSNKRSQAVGILMSSLHLDMDKIQHVVNLDNS VVDLETQALYENRAQSDLEKIEKHGRSSKDKENAKSLDKPEQFLYELSLIPNFSERVFCIL FQSTFSESICSIIRKLELLQKLCCTLKNGPGVMQVLGLVLAFGNYMNGNKTGQADGFGDLIL PKLKDVKSSDNSRSLLSYIVSYLLRNFDDEAGKEQCLFPLPEPQDLFQASQMKFEDFQKDLRK LKKDLKACEVEAGKVYQVSSKEHMQPFKENMEQFIQAKIDQEAENSLTETHKCFLETTAYF FMKPKLGEKEVSPNAFFSIWHEFSSDFKDFWKKENKLLQERVKEAEVCRQKKGKSLYKIKP RHDGKIKAKISMKT</p>

Further analysis of the NOV67a protein yielded the following properties shown in Table 67B.

Table 67B. Protein Sequence Properties NOV67a	
PSort analysis:	0.7000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV67a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 67C.

Table 67C. Geneseq Results for NOV67a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV67a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU97678	Human formin protein - Homo sapiens, 235 aa. [CN1294195-A, 09-MAY-2001]	1292..1526 1..235	235/235 (100%) 235/235 (100%)	e-135
ABB59865	Drosophila melanogaster polypeptide SEQ ID NO 6387 - Drosophila melanogaster, 1059 aa. [WO200171042-A2, 27-SEP-2001]	940..1525 480..1055	231/629 (36%) 317/629 (49%)	1e-99
AAB99983	Human limb deformation protein fragment - Homo sapiens, 199 aa. [CN1281047-A, 24-JAN-2001]	1355..1526 28..199	169/172 (98%) 171/172 (99%)	3e-95
AAB99982	Human limb deformation protein SEQ ID NO: 7 - Homo sapiens, 199 aa. [CN1281047-A, 24-JAN-2001]	1355..1526 28..199	169/172 (98%) 171/172 (99%)	3e-95
AAW76733	Mouse mDia Rho targeting protein - Mus sp, 1255 aa. [JP10262680-A, 06-OCT-1998]	887..1503 569..1149	233/640 (36%) 325/640 (50%)	1e-86

- 10 In a BLAST search of public sequence databases, the NOV67a protein was found to have homology to the proteins shown in the BLASTP data in Table 67D.

Table 67D. Public BLASTP Results for NOV67a				
Protein Accession Number	Protein/Organism/Length	NOV67a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9JL04	Formin 2 - Mus musculus (Mouse), 1567 aa.	1..1526 1..1567	1268/1594 (79%) 1328/1594 (82%)	0.0
Q9NZ56	Formin 2 - Homo sapiens (Human), 632 aa (fragment).	352..678 1..332	323/335 (96%) 324/335 (96%)	0.0
Q05858	Formin (Limb deformity protein) - Gallus gallus (Chicken), 1213 aa.	940..1525 647..1206	313/595 (52%) 403/595 (67%)	e-165
A41724	limb deformity (ld) protein - chicken, 1213 aa.	940..1525 647..1206	312/595 (52%) 402/595 (67%)	e-165
Q05859	Formin 1 isoform IV (Limb deformity protein) - Mus musculus (Mouse), 1206 aa.	955..1525 642..1199	296/581 (50%) 383/581 (64%)	e-156

PFam analysis predicts that the NOV67a protein contains the domains shown in the Table 67E.

Table 67E. Domain Analysis of NOV67a			
Pfam Domain	NOV67a Match Region	Identities/ Similarities for the Matched Region	Expect Value
FH2	1087..1491	126/506 (25%) 318/506 (63%)	5.7e-76

5

Example 68.

The NOV68 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 68A.

Table 68A. NOV68 Sequence Analysis			
	SEQ ID NO: 247	5476 bp	
NOV68a, CG138362-01 DNA Sequence	GCGGGAGCCCGCAGCGGAGGAGGGCCGCGGGCGGAGGCGCTGCGAGGCGCTGCTGGCCCTC GGGCTGCGGGAGCCGGCTAGGACGCCGAGCGGTGACTCCCCGCCGCTCCCGGGGGCCGTGAC GAAAGCAAACTATCTCTCAATATACCTCAGAAACAAAGATGTCTCCATCAAGTTTATACTCA CAGCAAGTGCTATGTTCTTCAATACCTTTATCGAAAAATGTGCACAGTTTTTTCAGTGCCTTC TGCACAGAAGATAATATTGAACAGAGTATCTCATATCTTGATCAGGAATTGACTACTTTTGGT TTTCCTTCATTATATGAAGAATCCAAAGGTAAAGAGACAAAGAGAGAGTTAAATATAGTAGCT GTACTAAATTGTATGAATGAGCTGCTTGTGCTTCAGCGGAAGAACCTTCTAGCTCAGGAAAAT GTGGAGACACAGAATTTGAAGCTGGGAAGTGATATGGACCATCTACAGAGCTGCTACTCAAAA CTTAAGGAACAACTGGAAACCTCAGGAGGGAATGATTGGGCTTCAGGAAAGAGACAGACAG TTACAATGTAAGAACAGGAATTTGCATCAGCTACTAAAGAATGAGAAAGATGAGGTGCAAAA TTACAAAATATCATTGCAAGTCGAGCTACTCAGTATAATCATGATATGAAGAGAAAAGAGCGT		

GAATATAATAAACTGAAGGAACGTCTACATCAACTTGTATGAACAAGAAAGATAAGAAAAATA
GCTATGGACATTTTGAATTATGTCTGGGAGAGCTGATGGAAAAAGAGGCTCCTGGAGGACTGGT
AAAACCTGAAGCCAGGAATGAAGATGAATGTATAAAATCTCTTGAATGATTATGAATATCGT
CAGAAACAAATCCTAATGGAATGAGAACTTAAGAAGGTTCTTCAACAAATGAAAAAGGAA
ATGATTTCTCTCTTTCTCCCAAAAGAAGAAACCTAGAGAAAGAGTAGATGATAGTACAGGA
ACTGTTATTTCCGATGTTGAAGAAGATGCCGGGGAACCTAAGCAGAGAGAGTATGTGGGACCTT
TCCTGTGAACTGTGAGAGAGCAGCTTACAAACAGCATCAGAAAAAGTGGAGAATTTTGA
AGTCATGTAGAAAAGCTTGATAACCAAGTTTCAAAGGTACACCTGGAAGGTTTTATGATGAA
GATGTAATCTCAGCACAAGACCATGAACAAGAACTGAAAACTCGAGTTAGAAATTCAGCAG
TGTAAGAAATGATTAAAACTCAGCAACAGCTTTTACAGCAGCAGCTCGCTACTGCATATGAT
GATGATACCACCTTCACTATTACGAGACTGTTATTTGTGGAAGAAAGAACGCTCTCAAAGAA
GAATGGTCCCTTTTTAAAGAGCAGAAAAAGATTTTGAGAGGGAGAGACGAAGCTTACAGAA
GCCGCTATTCGCCTGGGATTGGAGAGAAAGGCATTTGAAGAAGAAAGAGCCAGTTGGTTAAAG
CAGCAGTTTCTAAATATGACTACCTTTGACCACCAGAACTCAGAAAAATGTGAACTTTTCAGT
GCCTTCTCAGGAAGTTCTGATTGGGACAATCTTATAGTGCACCTCGAGGCAACCGCAAAAGAA
CCTCACAGTGTCTAATGGGTCTCCAGTTTGCATGTCTAACTTACTAAATCTCTCTCTGCT
TCACCTTCCACTTCAGACTTTTGCCAGACAGCTTCTGTCATATCTGAACATAGTTCAATCAAT
GTACTGAATATACTGCTGAAGAAATTAACCAAAATCAGGTTGGAGGAGAAGCTACAAATCAA
AAATGGAGTGTGGCGTCAAGACCTGGATCAGGAAGGTTGCTATAGTGGATGCTCCTTGAGC
TACACAAATCTCATGTAGAAAAAGATGACTTACCTTAGACATGTGGACTGGAAATTTTTCAT
TTAATGTGTTTCATCAAGTTTACATCTAAGTTGAACAGGGTGTGTCATAAAGTCAGTTATCT
CTAATACTTAAGATGGTCTGAGTTGTTTGTGTTGGACTTCCCTGTCTTCCCCAAAGAGTTGA
AATCTTAAATCTATTTAAAGGATATAAAGCTTTGGATATGATTTTGTAGTAACAGAGCAT
CTGGTTCTGTGAATAAAGGAATGTATAGATGTTTGGATGGAACAAAGACACTAGACTGAGTT
TCCTCTTATAGGTATTAAAAATAGCACTTTTAGGAACTGATTATTGTAAATGTTTAAATTTG
TCTCAATATAGTTGGCATTGGAAGTTTAGCCTTTACTTGAATGTATAGTGTAGATTTTAAAC
AAAGCGAGTCTATATTTATTATGTTTAGTGTGGTTTGAATTACCTCTTTCATATGTTTAA
ATAAAGTGAATTTATGTATGTTTGTACATAGATACACATGATTATGTAAGAGGCTTTAAG
ATTTAAAGTTTTCACACAACCATAAGTATAGTATTTTCATGCCAGTAAATTTTGTAGTGGTA
TTCTGTTTACAGATGTATTAGGACCATGTATGCATTACATTTAAGAAATCTCTTTAATACATC
TGGGCAATAAATATTGAAGGTATTCCATGAAGCTGAGTTCTTTAGATAATCAACACTACTAA
CATTACATTTTGTAGATTTTATGACATTAGATTTTATTTGTATATGTAGAATATTATAAT
TTTTAAAGGACTATTGATGATAGAAGATAGGGGCAAGACGACAAAAGTACCTTTGAATAAA
ACAATTTAAGAAATGGTTTAAAGATATTGGATGATAGAAGATATTTAAGATATCTAGATGGTG
ATATTTTCTTACAAGATGGGTACAGTATAGTAATATCTGTATACTAACTAGGGCTTTGTAT
TGTCAATAATTTTAAATAATTTTAAATGAGGTATTTACCACTGAAGAAATATGATAATATA
AAACCATCAAAATTTTATAATTTAGATGATCTCTGGAACAAATGTCATTTTCAATTTTCAAGAA
ACTCTTAAGCTCTCTCAGTCTCTGTAATGTTTCTGATTGCATGTTCTTCTATGAAAAGTATG
TTGTTGTTTGTAGTAGTAATAATAAATGATAGGCTCAGCTCTTCCAGGATTTTCATCAAA
AAGCTTTAAGTGCCTAACCTGCTGTCTCTGTACATAGAAGCCTGCACAGATCCAACCTTG
CTAGGTATCATAGTTTAGGCCATTTACCTTCCCTGTACTGGCAGTTCAGCCGCTTACATGC
ACTCACCTGTTTGTGGCTATTTTAAATTCATATTATTAACCAACCAAAACCAACCTATT
TGTGTTGTCCATTCACTATACGTAACATGTAACCTTTGGAGTTGCATAGCAGCAGCCATTT
TTTCAGGGCTGATAGGATATCATTAAGAGTGTCTTATGAGACATTAGTGGATATACTCATAGT
AACCATATTTATAGTTTAAAGAGCTAGCTCTTGAGCATTAGTCACTACCTTCAGCTTGATGC
ATGGTCACTTCTTACTATTTAAATACTACACATTGTACAAAATATCGAAGACACTACCATTA
TGCTAAAGGAAGAAATCTAGCTGGGATATAGGATTTTGTGTTTGTGTTTGTGTTTGTGTTTGT
CTATTTAGCAATAACATGGTCAAAAGGCAATCAGAAATTTAAATACAGTTAATGGATACAT
TTGGCAACAATTTCCCCGAGGGTTTTCCATGGTGTACTTTGCAAGAAATAGCACTCTAAT
TTTTAAAGTAAATCTCTTATTTTAGCAATATTATTTTCATGACAATGGAGTTCTAGAAAGC
AGCATCTGTTTTTGTGCGTGTTCATTTTAAAGTGAGTTGAGTTCTCAAATGGAAAGAA
AGATTCTTGTAGACGTACTTTTAAATCTAAAGTGTAAGAAACAGCAGAGTAAAGCCAGA
CTCATTCACCTTCAATGCTGTGATAGATCCAGAAGTTGTACATTTACCTAACACATCACT
TTGTTGAACATTCCTCACTCCAGAAATGATCCCCAATCACCTAATCTCAGAACTGCTGGAATGA
TGTCTGTTGGAACCCAGGACTCCACACACAAACCTCTGGGATTTTGTGTTTCCCTCTCTTT
CTAGGTGTTTGAATGTACAAATAATACAGCTGTGCTAATCTCACATTTAGCCATGATAGATG
ATGGTTCTAGAGTGTACTTCCATTTGTAAAGTCCCTGTAAAGTGTCTTCTGTTTATCACTA
TGTAATCTGAAATATTTGTACTTCATTTGTTTATCCATTCTGAATTTCAAAGCATAAA
AATGTCAAAGAAAAATGAGATAAACATTTGTTTACCATTAAATGTTTGCATCTGCAGTCTA
TGATGAGAAACATGAATATGACTAAAATTTAACAAATAGTATTTTCTTATAGAGCAATG
TATTAATGATTAGATACACAGGACAGTTCACAGCAAATGTTTACTTACATGGCATAATTGA
AATACGTATTATTGTAAAGAGTTCTATACTGCCTACCTTAGCTAACCAAGATTTGTCATAGAC
TTGTGATAATATGGTGTATTGGGAAACACATTTGTGCGCATGATGAAGAGAGCACCGTGAA
TCTAGTCTTGACTGCCCTGGTGGAAACATTTGAAAGCCCTTGGGACTCCAGGAGAGATGTTT
TGCAGCATGTATGCTAAATTTGTGAGTTAGCAAGACCCTGCTGTGATTACTATGTTGTACAGA
TCTTCAAAGTATATTGCCTAATATGAGTCATGTTATTTTAGATTGAATTTAAAGTAAGTATG

	GATTATGGAAGTCTAGTAAATATACTTTCCCTATTTTGTCTTTTTCAGTCTTTTGGTAAG CATTATTAAAGTACCTGCTGTATGTCAAGGCTTACTACTGCTTTCCAATCCTTGGTGATATGA AGACAGATAAGCACGCTGCCTGCTATTAGGAATCTGAGCTGAGTGAAGGCCAACTATTAAC CTACATTGTAATAAACAATGGTGGATGCTACGGACACATTCACTCTGTTTCATAAAACATGGAC TGATGGGTCTGCTCCAGGCCACATTTGTTTTTAACAGTAACGTGGCTAGGCTTCCATTTATA AGTCTTAGCATTATTTCTTTTGTGAAGTACTATGTAACAGATGATTGTTTGTAGATTTTGT TTCTACAATCAAAATGTTGACCTGCAAAGCAGTGTAGGATTTTCTCTCCTCAAGAGCGTGTAT TATTATCTAGTAGAAAAAGCATTCCCAAAGACTTGGTCCATGCAGATAAGGATAATGAAATTG CTCACTCTAATCCTTTTCTAAATACAGTGTTTTCAAGCTGGATGTAAATTAGAGTGGTGG ATATTGATTAAATTATTTGATTTATATGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA		
	ORF Start: ATG at 166		ORF Stop: TAG at 1927
	SEQ ID NO: 248	587 aa	MW at 68326.4kD
NOV68a, CG138362-01 Protein Sequence	MSPSSLYSQVLCSSIPLSKNVHSFFSAFCTEDNIEQSI SYLDQELTTFGFPSPLYEESKGKET KRELNI VAVLNCMNELLVLQRKNLLAQENVETQNLKLGSDMDHLQSCYSKLKEQLETSRREMI GLQERDRQLQCKNRNLHQLLNEKDEVQKLQNI IASRATQYNHDMKRKEREYNKLERLHQLV MNKKDKKIAMDILNYVGRADGKRGSWRTGKTEARNEDEMYKILLNDYEYRQKQILMENAELKK VLQMKKEMISLLSPQKKKPRERVDSTGTVISDVEEDAGELSRESMWDLSCTVREQLTNSI RKQWRILKSHVEKLDNQSVKHLEGFNDEDVISRQDHEQETEKLELEIQCKEMI KTOQQLLO QQLATAYDDTTSLRDCYLLEEKERLKEEWSLFKEQKKNFERERRSFTEAAIRLGLERKAFE EERASWLKQFLNMTTFDHQNSENVKLFSAFSGSSDWDNLIVHSRQPKKPHSVNSGSPVCMS KLTKSLPASPSTSDFCQTRSCISEHSSINVLNITABEIKPNQVGERTNQKWSVASRPGSQEG CYGCSLSYTNHVEKDDL		

Further analysis of the NOV68a protein yielded the following properties shown in Table 68B.

Table 68B. Protein Sequence Properties NOV68a	
PSort analysis:	0.5500 probability located in endoplasmic reticulum (membrane); 0.1900 probability located in lysosome (lumen); 0.1800 probability located in nucleus; 0.1000 probability located in endoplasmic reticulum (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV68a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 68C.

Table 68C. Geneseq Results for NOV68a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV68a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93250	Human protein sequence SEQ ID NO:12265 - Homo sapiens, 417 aa. [EP1074617-A2, 07-FEB-2001]	171..587 1..417	415/417 (99%) 415/417 (99%)	0.0
AAG03000	Human secreted protein, SEQ ID NO: 7081 - Homo sapiens, 137 aa. [EP1033401-A2, 06-SEP-2000]	245..381 1..137	137/137 (100%) 137/137 (100%)	2e-72

AAG40470	Arabidopsis thaliana protein fragment SEQ ID NO: 50219 - Arabidopsis thaliana, 382 aa. [EP1033405-A2, 06- SEP-2000]	29..381 28..361	89/354 (25%) 182/354 (51%)	2e-29
AAG48364	Arabidopsis thaliana protein fragment SEQ ID NO: 61066 - Arabidopsis thaliana, 347 aa. [EP1033405-A2, 06- SEP-2000]	18..336 13..339	83/335 (24%) 166/335 (48%)	2e-26
AAG33165	Zea mays protein fragment SEQ ID NO: 40144 - Zea mays subsp. mays, 291 aa. [EP1033405-A2, 06-SEP- 2000]	102..379 3..266	69/280 (24%) 148/280 (52%)	2e-22

In a BLAST search of public sequence databases, the NOV68a protein was found to have homology to the proteins shown in the BLASTP data in Table 68D.

Table 68D. Public BLASTP Results for NOV68a				
Protein Accession Number	Protein/Organism/Length	NOV68a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UIX0	Hypothetical 71.3 kDa protein - Homo sapiens (Human), 614 aa.	1..587 28..614	587/587 (100%) 587/587 (100%)	0.0
Q9Y2D8	KIAA0923 protein - Homo sapiens (Human), 614 aa.	1..587 28..614	586/587 (99%) 586/587 (99%)	0.0
Q8VC66	Hypothetical 71.0 kDa protein - Mus musculus (Mouse), 615 aa.	1..587 28..615	517/588 (87%) 549/588 (92%)	0.0
AAH31527	Expressed sequence AU014939 - Mus musculus (Mouse), 615 aa.	1..587 28..615	516/588 (87%) 549/588 (92%)	0.0
Q9FIE0	Genomic DNA, chromosome 5, P1 clone:MSF19 - Arabidopsis thaliana (Mouse-ear cress), 276 aa.	29..266 12..242	68/239 (28%) 132/239 (54%)	3e-25

5

PFam analysis predicts that the NOV68a protein contains the domains shown in the Table 68E.

Table 68E. Domain Analysis of NOV68a			
Pfam Domain	NOV68a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 69.

The NOV69 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 69A.

Table 69A. NOV69 Sequence Analysis			
	SEQ ID NO: 249	2254 bp	
NOV69a, CG138452-01 DNA Sequence	CTGGACTGGCCATTATGGAGGAGGAGCTGCAGCATTCCCATTGTGTGAATTGTGTCAGTAGAC GGTGCATGACCAGGCCAGAGCCAGGGATTTCCCTGTGATTGATTGGTTGTCCATTGGTTTGTG GTGCAGTTTTCATTCTTGTAAAGCTGATGAGCATCGACTTTTATGTCCATTGTAACGAGTGC CTTGCTTAAATAGTGACTTTGGATGTCCATTTACCATGGCCGAAATAAAGTTGCTGAACATC TAGAAATGTGCTCCTGCAAGTGTGGTGTGCTGTACTATGGAATGGAATCGATGGCCAGTTAGTT ATGCAGACCGGAAATCATATGAAATCTAAGCAGAGATGTCGATGAAGTGGCACAATTGGATA TGGCCTTGGCTCTTCAAGACCAAAGGATGCTCTTAGAATCCCTCAAAGTAGCCACCATGATGT CAAAAGCAACTGATAAAGTATCCAAACCTAGAGAACAATCTCAGTTAAATCAAGTGTCCCAG AAATACCACATGCTAATGGTTAGTGTCTGTGTGATGAAGAATCTTATGGTGCACCTTTATCAAG CTACTGTAGAAACAACCAGAAGTTTGGCTGCTGCTTTGGATATCCTGAATACTGCTACAAGAG ACATTGGCATGTTAAATACAAGTGTCCCAAATGACATGGATGAACAGCAAAATGCGAGAGAAA GCTTAGAGGATCAAACTTGAAAGACCAAGATCATCTTATGAGGAGGAAATAGGAGCAGTAG GTGGAATTGACTACAATGACACAAATCAGAATGCCAGTCTGAACAAATGGTTCAAGTGATT TATTATGTGACTTGAATACAAGTTCTTATGACACTTCTGCTCTTGTAAATGGCTTTCCTTTGG AAAATATATGTACCCAGGTCTAGACCAGAATCAGAATTTACATGGTGATTCAAAACAAAGTA ACTTAACAAATGGAGACTGTGTGGCATCATCAGATGGCACTTCAAACCTTCCAGCTCACTTG CGGTGGCAGCACAACTTAGGGAATAATACCATCCAGTGCTTTGCCTAATGGCAGAGTTCAGC ATATCCTCATGCCAGATGATGAAGGTGAAGGTGAATTGTGTTGGAATAAGTAGACTTAGGGG ACGTGAAGAATGTGGATGTCTTATCTTTCAGTCATGCTCCTTCATTCAATTTCTTCTTAATT CATGTTGGTCTAAACCAAAGGAAGATAAAGCAGTAGATACATCAGATTTGGAAGTTCAGAAAG ATCCTATGGGCCTCCAAGGAATAGATCTGATCACAGCAGCATTGCTTTTTTGTCTAGGAGATT CTCCAGGAGGGAGGGGTATATCTGATAGCCGCATGGCTGATATTATCACATTGACGTTGGGA CTCAGACTTTTTCACCTCCATCTGCAATATTAGCTACAAGTACAATGGTTGGGAGATAGCTT CAGCTTCAGCTTGTGATCATGCCAATCCACAGCTTTCAAATCCAAGTCCGTTTCAGACACTTG GGCTGGATTAGTATTGGAATGTGTCGCTAGGTACCAACCAAGCAGCGTTCAATGTTTACCT TTGTGTGTGACAGTTATTTAGAAGGAAAGAATTTTCTCCACTTTAAGAATGTGCATGGTG ACATTCATGCTGGACTCAATGGCTGGATGGAACAGAGGTGCCCTTTAGCTTACTATGGTTGTA CCTATTCTCAGCGTAGATTTTGTCCATCAATACAAGGAGCAAAGATTATACATGACCGCCATT TGAGGTCAATTGGAGTTCAAGCATGTGTATCTACAGTATTAGTGGAGCTGTAGAAACTGTG TGTGGGATTACATAATGACCATCTAAGTAGTCTTCTTTTGGAGTCTGCAGCATATTGCAG GCTTTCTCGATGGCTTCACTTATGTGAGTCTCATGTGTATCCAAGTTAATGAGGATGTGT GTGGCAGCCTGCTTCAGTCTCGTGGCATGGTCATACTGCAGTGGGGGAAAGGAAGTATCCAG AAGGAAATTCATCATGGCAGATAAAGAAAAGGTATGGCGATTAGTACTGCATTTTGTCTGTG TTAATGAATGAAATTTGCTGACATCTAAGCATGGCAGACCACTTGAAGAAATGCAGTTACA ATGTTGTGAGAAACGGGAGGAAGCAATCCCTTTGCCATGTATGTGTGTGACACGAGAACTCA CTAAGAAGGACGTTCACTACGCTCAGTTTAAACCTGTACTTTAAAA		
	ORF Start: ATG at 15		ORF Stop: TAA at 2250
	SEQ ID NO: 250	745 aa	MW at 82303.3kD
NOV69a, CG138452-01 Protein Sequence	MEEELQHSVCVNCVSRRCMTRPEPGISCDLIGCPLVCGAVFHSCKADEHRLLCFFERVPLNS DFGCPFTMARNVKVAEHLMECPASVVCCTMEWNRWPVSYADRKSYENLSRDVDEVAQLDMALAL QDQRMLESLSKVATMMSKATDKVSKPREQISVKSSVPEIPHANGLVSVDEESYGALYQATVET TRSLAAALDILNTATRDIGMLNTSPNDMDEQQNARESLDQNLKDDHLYEEIEIGAVGGIDY NDTNQNAQSEQNQGSSDLLCDLNTSSYDTSALCNGFPLENICTQVIDQNQLHGDGSKQSNLTNG DCVASSDGTSKPSSSLAVAAQLREIIPSSALPNGTVQHILMPDDEGEGLCWKKVDLGDVKNV DVLSFSHAPSFNLSNSCWSKPKEDKAVDTSDEVAEDPMGLQGIDLITAALLFCLGDSPPGR GISDSRMADIYHIDVGTQTFSLPSAILATSTMVGEIASASACDHANPQLSNPSPFQTLGLDLV LECVARYQPKQSRMFTFVCGQLFRRKEFSSHFKNVHGDHAGLNGWMEQRCPLAYYGCYSQR RFCPSIQGAKIIHDRHLRSFGVQPCVSTVLVEPARNCVLGLHNDHLSLPPFEVLQHIAGFLDG FSLCQLSCVSKLMRDVCGSLLSQSGMVILQWGRKYPEGNSSWQIEKVVWFSTAFCSVNEWK FADILSMADHLKKCSYNVVEKREEAIPLCMCVTRRELKEGRSLRSVLKPVL		

Further analysis of the NOV69a protein yielded the following properties shown in Table 69B.

Table 69B. Protein Sequence Properties NOV69a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

- 5 A search of the NOV69a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 69C.

Table 69C. Geneseq Results for NOV69a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV69a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU14823	Novel bone marrow polypeptide #29 - Homo sapiens, 415 aa. [WO200155442-A2, 02-AUG-2001]	133..515 21..403	383/383 (100%) 383/383 (100%)	0.0
AAU14910	Novel bone marrow polypeptide #116 - Homo sapiens, 101 aa. [WO200155442-A2, 02-AUG-2001]	658..745 14..101	88/88 (100%) 88/88 (100%)	5e-47
AAM23608	Human EST encoded protein SEQ ID NO: 1133 - Homo sapiens, 86 aa. [WO200154477-A2, 02-AUG-2001]	666..733 2..69	23/68 (33%) 39/68 (56%)	6e-08
ABB69873	Drosophila melanogaster polypeptide SEQ ID NO 36411 - Drosophila melanogaster, 3232 aa. [WO200171042-A2, 27-SEP-2001]	188..329 1527..1678	38/157 (24%) 68/157 (43%)	0.37
ABB71150	Drosophila melanogaster polypeptide SEQ ID NO 40242 - Drosophila melanogaster, 2858 aa. [WO200171042-A2, 27-SEP-2001]	134..345 758..944	46/217 (21%) 84/217 (38%)	1.4

- 10 In a BLAST search of public sequence databases, the NOV69a protein was found to have homology to the proteins shown in the BLASTP data in Table 69D.

Table 69D. Public BLASTP Results for NOV69a				
Protein Accession Number	Protein/Organism/Length	NOV69a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8TB52	Similar to RIKEN cDNA 1700026A16 gene - Homo sapiens (Human). 745 aa.	1..745 1..745	745/745 (100%) 745/745 (100%)	0.0
Q9D9X5	1700026A16Rik protein - Mus musculus (Mouse), 746 aa.	1..745 1..746	653/749 (87%) 696/749 (92%)	0.0
Q9BXZ7	F-box domain protein - Homo sapiens (Human), 390 aa.	356..745 1..390	388/390 (99%) 389/390 (99%)	0.0
Q9UH90	Muscle disease-related protein - Homo sapiens (Human), 709 aa.	7..733 12..693	254/745 (34%) 376/745 (50%)	e-111
Q9ULM5	KIAA1195 protein - Homo sapiens (Human), 717 aa (fragment).	7..733 20..701	251/745 (33%) 375/745 (49%)	e-109

Pfam analysis predicts that the NOV69a protein contains the domains shown in the Table 69E.

Table 69E. Domain Analysis of NOV69a			
Pfam Domain	NOV69a Match Region	Identities/ Similarities for the Matched Region	Expect Value
F-box	611..658	15/48 (31%) 34/48 (71%)	1.5e-05

5

Example 70.

The NOV70 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 70A.

Table 70A. NOV70 Sequence Analysis			
	SEQ ID NO: 251	753 bp	
NOV70a, CG138781-01 DNA Sequence	ATGCAGGACCCCAACGCAGACACTGAGTGGAAATGACATCTTACGCAAAAAGGGTTTCTTACCC GCCAAGGAAGAATTGGAAGAATTGGAAGAGGAGGCAGAAGAGGAGCAGCGCATCCTCCAGCAG TCAGTGGTGAAAACATATGAAGATATGACTTTGGAAGAGCTGGAGGATCACGAAGGCGAGTTT AATGAGGAGGATGAATGTGCTATTGAAATGTACAGACAGCAGAGACTGGCTGAGTGGAAAGCA ACTAACTGAAGAATAAATCTGGAAGTTTTGGAGATCTCAGGGAAGGATTATGTTCAAGAA GTTACCAAAGCTGGCGAGGGCTTGTGGTCATCTGCACCTTTACAAACAAGGAATTCCCCCTC TGTGCCCTGATAAATCAGCACCTCAGTGGACTTGCCAGGAAGTTTCTGTATGTCAAATTTATC AAAGCCATTTCAACAACCTGCATACCCAATTATACTGATAGGAATCTGCCCACGATATTTGTT TACCTGGAAGGAGATATCAAGGCTCAGTTATTGGTCCTCTGGTGTGTTGGCGGCATGAACCTG ACAAGAGATGAGTTGGAATGGAACCTGTCTGAATCTGGAGCAATTACGACAGACCTGGAGGAA AACCTAAGAAGCCGATTGAAGACGTGTTGCTCTCCTCAGTGCGGCGCTCTGTCTCATGAAG		

	AGGGACAGCGATTCCAAGGGTGACTGAGGCTACAGCTGCTATCCCATGCCGAACCTTCTT		
	ORF Start: ATG at 1		ORF Stop: TGA at 718
	SEQ ID NO: 252	239 aa	MW at 27409.8kD
NOV70a, CG138781-01 Protein Sequence	MQDPNADTEWNDILRKKGFLPAKEELEEEEEEAEQRI LQQSVVKTYEDMTLEELEDHEGEF NEEDECAIEMYRQQLAEWKATKLKNSGKVLEISGKDYVQEVTKAGEGLWVILHLYKQGIPL CALINQHLSGLARKFPDVKFIKAISTCIPNYTDRNLPTIFVYLEGDIKAQFIGPLVFGGMNL TRDELEWKLSGSAITTDLEENPKPIEDVLLSSVRRSVLMKRSDSDSKGD		

Further analysis of the NOV70a protein yielded the following properties shown in Table 70B.

Table 70B. Protein Sequence Properties NOV70a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV70a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 70C.

Table 70C. Geneseq Results for NOV70a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV70a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB68507	Human GTP-binding associated protein #7 - Homo sapiens, 239 aa. [WO200105970-A2, 25-JAN-2001]	1..239 1..239	226/239 (94%) 231/239 (96%)	e-128
AAE02001	Human viral IAP-associated factor (VIAF) - Homo sapiens, 239 aa. [WO200134798-A1, 17-MAY-2001]	1..239 1..239	226/239 (94%) 231/239 (96%)	e-128
AAU27979	Mouse contig polypeptide sequence #132 - Mus musculus, 243 aa. [WO200164834-A2, 07-SEP-2001]	1..239 5..243	226/239 (94%) 231/239 (96%)	e-128
AAU27807	Human full-length polypeptide sequence #132 - Mus musculus, 239 aa. [WO200164834-A2, 07-SEP-2001]	1..239 1..239	226/239 (94%) 231/239 (96%)	e-128
AAB43903	Human cancer associated protein sequence SEQ ID NO:1348 - Homo sapiens, 243 aa. [WO200055350-A1, 21-SEP-2000]	1..239 5..243	226/239 (94%) 231/239 (96%)	e-128

In a BLAST search of public sequence databases, the NOV70a protein was found to have homology to the proteins shown in the BLASTP data in Table 70D.

Table 70D. Public BLASTP Results for NOV70a				
Protein Accession Number	Protein/Organism/Length	NOV70a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9H2J4	HTPHLP (Unknown) (Protein for MGC:3062) - Homo sapiens (Human), 239 aa.	1..239 1..239	226/239 (94%) 231/239 (96%)	e-128
CAC40344	Sequence 3 from Patent WO0134798 - Mus musculus (Mouse), 240 aa.	1..239 1..240	207/240 (86%) 223/240 (92%)	e-116
Q99JX2	Similar to RIKEN cDNA 1110061A19 gene - Mus musculus (Mouse), 240 aa.	1..239 1..240	206/240 (85%) 222/240 (91%)	e-115
Q9D0W3	1110061A19Rik protein - Mus musculus (Mouse), 239 aa.	1..239 1..239	207/240 (86%) 222/240 (92%)	e-114
CAC40345	Sequence 5 from Patent WO0134798 - Brachydanio rerio (Zebrafish) (Zebra danio), 239 aa.	1..239 1..237	170/239 (71%) 203/239 (84%)	2e-97

5

Pfam analysis predicts that the NOV70a protein contains the domains shown in the Table 70E.

Table 70E. Domain Analysis of NOV70a			
Pfam Domain	NOV70a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Phosducin	64..179	33/121 (27%) 58/121 (48%)	0.36

10

Example 71.

The NOV71 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 71A.

Table 71A. NOV71 Sequence Analysis			
	SEQ ID NO: 253	6765 bp	
NOV71a,	GTTAAATGGGCAACTCCGACAGTCAGTACACCCTTCAAGGATCTAAAAATCATAGCAATACT		

CG138808-01 DNA
Sequence

ATTACTGGTGCTAAGCAAATTCCTTGCTCCCTGAAAATACGTGGCATTTCATGCAAAAGAGGAA
AAGTCATTGCATGGATGGGGTCACGGAAGCAACGGAGCAGGTTACAAGTCCAGGTCCTGGCC
CGAAGCTGCCTTTCTCACTTTAAGAGTAACAGCCTTACGCATCGAGACTCGGTGGCCCCACA
TGCAAGGTCTCCAGAGGTGTTGCCACTCCACGCACAGGACAAATGCCCCAGGGAAGGATTTT
CAGGGCATCAGTGCTGCTTTCTCAACTGAGAATGGCTTCCACTCTGTTGGCCACGAGCTGGCA
GATAACACATCACCTCCAGAGACTGCAACGGACACCTTCTCAACTGTACGGGAGGAATGAG
AGCATTGCCTCCACCCACCGGGCGAAGACCGCAAGAGCCCCGAGTGCTCATCAAAACGCTG
GGGAAGCTGGATGGGTGTTAAGGGTCGAGTTCACAATGGTGGCAACCCAGCAAAGTGCCT
GCAGAGGACTGCAGTGAGCCGCTGCAGCTGCTGAGGTACTACCTTACCTTAGCATCGGAAACC
TCCCCTGTGCCTGAAGCCAGGAGGGGTCCAGCGCGGATTCCCCTGCCAGCCATCGCCCTCT
CCCACGGACTCTCGCTCGGTCCAGCAAAGGCAGCTCCCTGAGTTCGTAGTCATCTCTGGTAC
GACTCCCCTTGGGGCAATGCTGGAGAGCTGAGCGAGGCTGAGGGCTCCTTCTGGCCCCGGC
ATGCTGACCCAGTCTCCATGCCAGCTTCCACCTGGCGATGCCAAAAAGCCTTTCAACCAA
AGCTCTTCCCTCTCCTCCCTCCGGGAAGTGTACAAAGATGCCAACCTGGGGAGCCTCTCCCC
TCAGGTATCCGCTTTCTGATGAATACATGGGCACGCATGCCAGCTGAGCAACCGTGCTCT
TTTGCTTCCGACATTGATGTGCCCTCCAGAGTGGCACACGGGGACCCATCCAGTACAGTTC
TTCACTCTCCCTGTGGAAGCCCAAAGCCTTGTGAGGATATGCGAAGAAGGACTCCCTC
AAAGCCAGGATGCGACGGATCAGTGACTGGACGGGAAGCCTCTCAAGGAAGAAAAGGAACTC
CAGGAGCCGAGGTCCAAGGAGGCGAGTACTTGTGACAGTCGCTCTGATGGACTGAATACA
GATGTGCAGGATCTCCACGGCATCTGCTTTCTGTGGTCAGGGGGCTCTACTCAGATCCTG
TCTCAGAGAAGTGAATCCACACATGCGATTGGCAGCGATCCCCTCCGGCAGAACATTTATGAG
AATTCATGCGAGAGTTGGAATGAGCAGGACCAACACTGAGAACATAGAACATCTACAGAA
ACCGCCGAGTCCAGCAGCGAGTCACTCAGCTCTCTGGAACAGCTGGATCTGCTCTTTGAGAAG
GAACAGGGGGTGGTCCGGAAGGCCGGGTGGCTCTTCTCAAGCCCTGGTCACTGTGCAAG
GAAAGGAAGCTTGAGCTGGTGGCACGAAGGAAATGGAACAGTACTGGGTAACGCTGAAAGGT
TGCACGCTGCTGTTTATGAGACCTATGGGAAGAATCCATGGATCAGAGCAGTCCCCCTCGG
TGTGCTCTGTTTGCAGAAGACAGCATAGTGAGTCTGTTCCAGAGCATCCCAAGAAAGAAAT
GTGTTCTGCCTCAGCAACTCCTTTGGAGATGTCTACCTTTCCAGGCCACCAGCCAGACAGAT
CTAGAAAAGTGGGTCACTGTGTACACTCTGCTTGTGTCATCCCTTTTGCAGAAAGAGCATGGG
AAAGAGGACACGCTCGGCTGCTGAAGAACCAGACCAAAACCTGCTTCAGAAGATAGACATG
GACAGCAAGATGAAGAAGATGGCAGAGCTGCAGCTGTCCGTGGTGAGCGACCCAAAGAACAGG
AAAGCCATAGAGAACAGATCCAGCAATGGGAGCAGAATCTTGAGAAATTTACATGGATCTG
TTAGGATGCGCTGCTATCTGGCCAGCCTACAAGGTGGGGAGTTACCGAACCCAAAGAGTCTC
CTTGACGCCGCCAGCCGCCCTCCAAGCTGGCCCTCGGCAGGCTGGGCATCTTGCTGTTTCC
TCTTTCCATGCTCTGGTATGTTCTAGAGATGACTCTGCTCTCCGGAAGGACACTGTCACTG
ACCCAGCGAGGGAGAAACAAGAAGGAATATTTCTCTGTTAAAAGGGCTGGACACACTGGCC
AGAAAAGGCAAGGAGAAGAGACCTTATAACTCAGGTGTTTGAATCAAGTGGCAGCCATGGA
TTTTCTGGAAGTCACTACCTCAAACTCCAGTAAGTCCAGTGAGGTCGATGAATCTTGCAT
ATATATGGTTCAACAGTAGACGGTGTTCCTCCGAGACAATACATGGGAAATCCAGACTTATGTC
CACTTTCAGGACAATCACGGAGTACTGTAGGGATCAAGCCAGAGCAGAGTAGAAGATATT
TTGACTTTGGCATGCAAGATGAGGAGTGGAAACCCAGCCATTATGGCCTACAGCTTCGAAAA
TTAGTAGATGACAATGTTAGTATGTCATCCCTGCACCATATGAATATATGCAACACAGGTT
TATGATGAAATAGAAGTCTTTCCACTAAATGTTTATGACGTGCAGCTCACGAAGACTGGGAGT
GTGTGTGACTTTGGGTTTGCAGTTACAGCGCAGGTGGATGAGCGTCAGCATCTCAGCCGGATA
TTTATAAGCGACGTTCTTCCCGATGGCCTGGCGTATGGGGAAGGGCTGAGAAAGGGCAATGAG
ATCATGACCTTAAATGGGGAAGCTGTGTCTGATCTTGACCTTAAGCATGAGGAGCCCTGTTT
TCTGAGAAGAGCGTCGGAAGTCACTCTGATTGCCCGGCTCCGGACACAAAAGCAACCTGTGT
ACATCCTGGTCAGACAGTACCTGTTCTCCAGGGACAGAAAGTCTGCTGCCCTCTCTAAC
CAGTCCCAACTGCTGGAGGAATCTTGATAACTTTAAAAAGAAATACAGCCATGATTTACAGC
AACGTCCTGATATCACACAGGTCTGAAAGGAGTCAGACAGATGGCACTCTGGATCAGGTT
TCCCACAGGGAGAAAATGGAGCAGACATTAGGAGTGTGAGCAGATCACTGCACTGTGCAAG
AGTTTAAACGACAGTCAGGCCAACGGCATGGAAGGACCGGGGAGAAATCAGGATCCTCCTCG
AGGCCTCTGGCCCCCAGCTGTCTGATGCAGACCGCTCCGCAAGTATCCAGGAGCTTGTG
GACACAGAGAAGTCTACGTGAAGGATTTGAGCTGCCTCTTTGAATTATACTTGGAGCCACTT
CAGAATGAGACCTTTCTTACCAGATGAGATGGAGTCACTTTTGGAGGTTTGGCAGAGATG
CTTGAGTTTCAAGAGGTGTTTCTGGAGACCTGGAGGATGGGATTTAGCATCATCTGACTTT
AACACCTAGAAAACCCCTCACAGTTAGAAAATTAAGTCTTCCCTGGAGGCTCTTCTCTT
TATTACGCGGACCACTTTAACTGTACAGTGGATTCTGTGCTAACCATATCAAGTACAGAAG
GTTCTGGAGCGAGCTAAAAGTACAAAGCCTTCAAGGCTTTCTGGACGCCCGGAACCCACCC
AAGCAGCATTCCTCCACGCTGGAGTCTTACCTCATCAAGCCGGTTCAAGAGTGTCTCAAGTAC
CCGCTGCTGCTCAAGGAGCTGGTGTCCCTGACGGACAGGAGAGCGAGGAGCACTACCACCTG
ACGGAAGCACTAAAGGCAATGGAGAAAGTAGCGAGCCACATCAATGAGATGAGAAGATCTAT
GAGGATTATGGGACCGTGTGTGACCAGCTAGTAGCTGAGCAGAGCGGAACAGAGAAGGAGGT
ACAGAACTTTTGGATGGAGAGCTTCTGATGGCACTTACGGTTTCTGGTGTGAATCCATTCTG
TCTCTAGGAAAAGCTAGAAAGACCTTGAGCTCACAGTATTGTTTTTAAGAGAGCCGTCATA
CTGTTTATAAGAAAAGTGAAGTGAAGAAATGACCTCGAATCCCGGCTGCACAC

	A A C T C T A C T G A C T T G G A C C C A T T T A A A T T C C G C T G G T G A T C C C A T C T C C G C G T T C A A G T C A G A C T G G G G A A T C C A G C A G G G A C A G A A A T A A T T C C A T A T G G G A A C T G A T C C A T A C G A A G T C A G A A A T A G A A G G A C G G C C A G A A A C C A T C T T T C A G T T G T G T T G C A G T G A C A G T G A A G C A A A A C C A A C A T T G T T A A G G T G A T T C G T T C A T T T C T G A G G G A G A A C T T C A G G C G T C A C A T A A A G T G T G A A T T A C C A C T G G A G A A A C G T G T A A G G A T C G C C T G G T A C C T C T T A A G A A C C G A G T T C C T G T T T C G G C C A A A T T A G C T T C A T C C A G G T C T T T A A A A G T C C T G A A G A A T T C C T C C A G C A A C G A G T G G A C C G G T G A G A C T G G C A A G G G A A C C T T G C T G G A C T C T G A C G A G G G C A G C T T G A G C A G C G G C A C C C A G A G C A G C G G C T G C C C C A C G G C T G A G G G C A G G C A G G A C T C C A A G A G C A C T T C T C C C G G G A A A T A C C C A C A C C C G G C T T G C C A G A T T T T G C T G A C A A T C T C A T C A A G A G A G T G A C A T C C T G A G C G A T G A A G A T G A T G A C C A C C G T C A G A C T G T G A A G C A G G G C A G C C C T A C T A A A G A C A T C G A A A T T C A G T T C C A G A G A C T G A G G A T T T C C G A G G A C C C A G A C G T T C A C C C C G A G G C T G A G C A G C A G C T G G C C C G G A G T C G G G T G A G G G T C A G A A A G G A G A G A G C A G C C C A A A C T G G T C C G G G G C A C T T C T G C C C C A T T A A A C G A A A G C C A A C A G C A C C A A G A G G G A C A G A G G A A C T T T G C T C A A G G C G C A G A T C C G T C A C C A G T C C C T T G A C A G T C A G T C T G A A A T G C C A C C A T C G A C C T A A A T T C T G T T C T A G A C C G A G A A T T C A G T G T C C A G A G T T T A A C A T C T G T T G T C A G T G A G G A G T G T T T A T G A A A C A G A G A G C C A C G G A A A T C A T A G T A T G A T T C A A T C C A G A T A T G G G T T A A A T T C C T C A T T T T A C T T T T A A A C T G G T G T T A A G T G G A A A T T G C A A A A A A A A A A A A A A A A C T G T T C A T T C C T G G G T T T T G T G C A G T A T A C A T T T T C C C A C A A A T G G T T G T A A A G A T T T A A G T A T T T T A A T T T A T T G T G G A T C A G A A A C C T A G A T G A A A C T G G T C A G A A T C T G A A A T T A C T T A G T T T A T A T C C A C T T T G A G C A G G T A T C A A A T G A T T T A G G A T C C T T A A A A T T A C A T T C T A A A T T A A G T T A T G T G G A A A A G T A A G G C T G G G A A G T C G T G A T T A A T A G T T T T C A A A G G C C A T T T T T A A A A T C C T C T G G G C A T T T T C T T T C A G C T G T T T G T T A G T T T T T G C T T A T T T A A A G C A T A T T T A A G T A T T T T A A T G T G G T T T A G G G C A A A A T G T G C A G A T A C T T C A T T T T G T A A G A T A G A T T G T A A T A G A T G C T G T T A T A C T A A A C A T G T C A A C T A T C T A T A C A G T A T A T A T A A A A G A A A G C T T G T A C T G T A T C T A T T T T G A T A T A T T T A T T T C T C T G C C A A G C T G T A T A G T A A A A G G A A A T A A G T C A C A T C T G G T C A T T G G C A T T T G T A T C G T C A T T C T G T A A A G A C A A A A G A G T A C C T A T A T A A A A G C T C C A C G T A G T G C A A A T C G A C A T C T G G T A G G C T G C T C G C C C C A G G C A G C A G C T A G A G T C T G T A A T T C T C T G C G T C A T C C T C T C T T T T C T T C A T T T T T G C T T T T C T T C G C T T G A G T T C T C T C T G A A A T T A T A T G C A A A G A G T T G T G G G T C T T C A T C A C A C A T T T T T C T G T A T A C A T C A C A G A G G C T C T T A A G T G T G A G A T G G A G A G C T G T G G G G C C G A A G A G T A G G G T C T A T G T C T G C C A A C T C T A A C A G C C T C C C G T G C T T T C C A A G C G C T G C G C T T C A G G G A A T A A C A T T C T G A G C C C T C G A T G G C A G T A T T T C C T C G G A A C T G A A A T A C A T T C T G A A C C A C T T T T T C C A C C A G C T T G A A T G G C T G C T C T A T C A T T G C T G T A T C A A G G G A G T G A A G T C A C C A C G C C C A C G T C C A C C T T C G T T G T A A G C A A A C A T A T T A T C A T T C T G T G C A T G A T A T G T G C A T A G T G T G A T C A A T C A A C T C A T C C T T G T A A A A C A G G A A G A T G G G C T G T C A A C A G C C T G T T T C A T A A A C A G A C C T T T C C A C G T A C T T C G G T T T C A T C T C T A G G C A T G G A A G A T G G T A C A T T C T G A T T C G C A A T G A C A T G G A G A A A C A G C C G C T G C A C C T G T T C T C T A A T G A C A T C C A C C A G A C C T G T G C T T G A T G G T C A C T T A A T T T T A A A A C A C A G T T T C A A A T G G C T T A A A A A T C A A T C C A A A T C A G T A A A G T C A G T C A G C A G A T A A T A G A T G G C A T T A G A A T A T T T T A G T T T T T G A A T G A G G A A A A A A T A A G C T G C A G C A G C A G C T T C A A G A C A C A G A G A T G G C A G A C A G G C C C C C A G G G A C C A C T C A G T G C T A A A C T T C C C A G A T A G A G A C A C C A C T T A T T T T C G G T A G A C A C T G A T T A A T C A G T G G A C T G A A T T C		
	ORF Start: ATG at 7 ORF Stop: TAG at 5182		
	SEQ ID NO: 254	1725 aa	MW at 192579.1kD
NOV71a, CG138808-01 Protein Sequence	M G N S D S Q Y T L Q G S K N H S N T I T G A K I P C S L K I R G I H A K E E K S L H G W H G S N G A G Y K S R S L A R S C L S H F K S N Q P Y A S R L G G P T C K V S R G V A Y S T H R T N A P G K D F Q G I S A A F S T E N G F H S V G H E L A D N H I T S R D C N G H L L N C Y G R N E S I A S T P P G E D R K S P R V L I K T L G K L D G C L R V E F H N G G N P S K V P A E D C S E P V Q L L R Y S P T L A S E T S P V P E A R R G S S A D S L P S H R P S P T D S R L R S S K G S S L S S E S S W Y D S P W G N A G E L S E A E G S F L A P G M P D P S L H A S F P P G D A K P F N Q S S L S L R E L Y K D A N L G S L S P S G I R L S D E Y M G T H A S L S N R V S F A S D I D V P S R V A H G D P I Q Y S S F T L P C R K P K A F V E D T A K K D S L K A R M R R I S D W T G S L S R K K R K L Q E P R S K E G S D Y F D S R S D G L N T D V Q G S S Q A S A F L W S G G S T Q I L S Q R S E S T H A I G S D P L R Q N I Y E N F M R E L E M S R T N T E N I E T S T E T A E S S S E S L S S L E Q L D L L F E K E Q G V V R K A G W L F F K P L V T V Q K E R K L E L V A R R K W K Q Y W V T L K G C T L L F Y E T Y G K N S M D Q S S A P R C A L F A E D S I V Q S V P E H P K K E N V F C L S N S F G D V Y L F Q A T S Q T D L E N W V T A V H S A C A S L F A K K H G K D T L R L L K N Q T K N L L Q K I D M S K M K M A E L Q L S V S D P K N R K A I E N Q I Q Q W E Q N L E K F H M D L F R M R C Y L A S L Q G G E L P N P K S L L A A A S R P S K L A L G R L G I L S V S F H A L V C S R D D S A L R K R T L S L T Q R G R N K K G I F S S L K G L D T L A R K G E K R P S I T Q V F D S S G S H G F S G T Q L P Q N S S N S S E V D E L L H I Y G S T V D G V P R D N T W E I Q T Y V H F Q D N H G V T V G I K P E H R V E D I L T L A C K M R Q L E P S H Y G L Q R K L V D D N V E Y C I P A P Y E M Q Q Q V Y D E I E V F P L N V Y D V Q L T K T G S V C D F G F A V T A Q V D E R Q H L S R I F I S D V L P D G L A Y G E L R K G N E I M T L N G E A V S D L D L K Q M E A L F S E K S V G L T L I A R P P D T K R T L C T S W S D S D L F S R D Q S L L P P P N Q S Q L L E E F L D N F K N T A N D F S N V P D I T T G L K R S Q T D G T L D Q V S H R E K M E Q T F R S A E Q I T A L C R S F N D S Q A N G M E G P R E N Q D P P P R L A R H L S D A D R L R K V I Q E L V D T E K S Y V K D L S C L F E L Y L E P L Q N E T F L T Q D E M E S L F G S L P E M L E F Q K V F L E T L E D G I S A S S D F N Q L E T P S Q F R K L L F S L G G S F L Y Y A D H F K L Y S G F C A N H I K V Q K V L E R A K T D K A F A F L D A R N P T K H S T L E S Y L I K P Q R V L K Y P L L L K E L V S L T D Q E S E E H Y H L T E A L K A M E K V A S H I N E M Q K I Y E D Y G T V F D Q L V A E Q S G T E K E V T E L S M G E L L M H S T V S W L N P F L S L G K A R K D L E L T V F V F K R A V I L V		

YKENCKLKKKLPSNSRPAHNSTDLDPFKFRWLIPISALQVRLGNPAGTENNSIWELIHTKSEI EGRPETIFQLCCSDSESKTNIIVKVISILRENFRRIKCELPLEKTCKDRLVPLKNRVPVSAK LASSRSLKVLKNSSSNEWTGETGKGTLLDSDSEGLSSGTQSSGCPTAEGRODSKSTSPGKYPH PGLADFADNLIKESDILSDEDDHRTVKQGSPTKDIEIQFQRLRISEDPDVHPEAEQPGPE SGEGQKGGEQPKLVRGHFCPIKRKANSTKRDRGTLLKAQIRHQLSDSQSENATIDLNSVLERE FSVQSLTSVVSEECFYETESHGKS

Further analysis of the NOV71a protein yielded the following properties shown in Table 71B.

Table 71B. Protein Sequence Properties NOV71a	
PSort analysis:	0.9800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV71a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 71C.

Table 71C. Geneseq Results for NOV71a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV71a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB07497	A T-cell lymphoma invasion and metastasis 2 (TIAM2) protein - Homo sapiens, 1077 aa. [WO200040607-A2, 13-JUL-2000]	649..1725 1..1077	1075/1077 (99%) 1076/1077 (99%)	0.0
AAB07496	A T-cell lymphoma invasion and metastasis 2 (TIAM2) protein - Homo sapiens, 1053 aa. [WO200040607-A2, 13-JUL-2000]	649..1725 1..1053	1052/1077 (97%) 1052/1077 (97%)	0.0
AAU21607	Novel human neoplastic disease associated polypeptide #40 - Homo sapiens, 719 aa. [WO200155163-A1, 02-AUG-2001]	1007..1725 1..719	719/719 (100%) 719/719 (100%)	0.0
AAB07495	A T-cell lymphoma invasion and metastasis 2 (TIAM2) protein - Homo sapiens, 626 aa. [WO200040607-A2, 13-JUL-2000]	1100..1725 1..626	626/626 (100%) 626/626 (100%)	0.0

AAB07494	A T-cell lymphoma invasion and metastasis 2 (TIAM2) protein - Homo sapiens, 626 aa. [WO200040607-A2, 13-JUL-2000]	1100..1725 1..626	626/626 (100%) 626/626 (100%)	0.0
----------	---	----------------------	----------------------------------	-----

In a BLAST search of public sequence databases, the NOV71a protein was found to have homology to the proteins shown in the BLASTP data in Table 71D.

Table 71D. Public BLASTP Results for NOV71a				
Protein Accession Number	Protein/Organism/Length	NOV71a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9WVS3	Sif and Tiam1-like exchange factor - Mus musculus (Mouse), 1715 aa.	1..1725 1..1715	1456/1726 (84%) 1549/1726 (89%)	0.0
Q9UKW0	T-cell lymphoma invasion and metastasis 2 - Homo sapiens (Human), 1077 aa.	649..1725 1..1077	1075/1077 (99%) 1076/1077 (99%)	0.0
Q9UFG6	Hypothetical 118.0 kDa protein - Homo sapiens (Human), 1046 aa (fragment).	676..1725 21..1046	1024/1050 (97%) 1025/1050 (97%)	0.0
Q9UKV9	T-cell lymphoma invasion and metastasis 2 short - Homo sapiens (Human), 626 aa.	1100..1725 1..626	626/626 (100%) 626/626 (100%)	0.0
Q13009	T-lymphoma invasion and metastasis inducing protein 1 (TIAM1 protein) - Homo sapiens (Human), 1591 aa.	297..1623 230..1535	593/1353 (43%) 814/1353 (59%)	0.0

5

PFam analysis predicts that the NOV71a protein contains the domains shown in the Table 71E.

Table 71E. Domain Analysis of NOV71a			
Pfam Domain	NOV71a Match Region	Identities/ Similarities for the Matched Region	Expect Value
PH	507..620	26/114 (23%) 87/114 (76%)	1.8e-17
PDZ	914..999	22/89 (25%) 59/89 (66%)	4.8e-06
RhoGEF	1127..1316	70/211 (33%) 140/211 (66%)	3.3e-40

Example 72.

The NOV72 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 72A.

Table 72A. NOV72 Sequence Analysis	
	SEQ ID NO: 255
NOV72a, CG139224-01 DNA Sequence	4172 bp
	<p>CGGGCCGCCCCGGCTAAGAGCGGCCGGCTGGAGCCGCTGAGCCCCCGCTGCGGCCGGGAGCTC CATGGGGAGCGCCGCGCAGCGCTTGGGAAGATGCCCCGGGAGCTGCCCTGCCGAGGGG TGGGAGAGGCGCGCACTTCAGCGGCAAGGCTCTACTACATGAGACCACGAACCGCACCA AGCTGGATCGACCCGCGGGACAGGTACACCAAACCGCTCACCTTTGCTGACTGCATTAGTGAT GAGTTGCCGCTAGGATGGGAAGAGGCATATGACCCACAGGTTGGAGATTACTTCATAGACCAG AACACCAAAACCACTCAGATTGAGGATCCTCGAGTACAATGGCGGGGAGCAGGAACATATC CTGAAGGATTACTGGTGGTGGCCAGGAGGCTCTGAGTGCACAAAGGAGATCTACACAGGT AAGCAGCAGCGCTGGAGCTTGACACAGCAGGAGTACCAGCAACTGCATGCCGCTCTGGAGCAT AAGCTGGGCTCCCAGGTCAGCGTGGTCTCTGGTTTCATCATCCAGCTCCAAGTATGACCCTGAC ATCCTGAAAGCTGAAATTGCCACTGCAAAATCCCGGGTAAACAAGCTGAAGAGAGAGATGGTT CACCTCCAGCAGAGCTGCAGTTCAAGAGCGTGGCTTTACAGACCTGAAGAAGATCGATAAG AAAATGCTGATGCTCAGGGCAGCTACAAACTGAGTGAAGCTCAGAGTCTGTGAGAGAAACA AAAGCCATCAAAAAGGCTATTACCTGTGGGAAAAGGAAAAGCAAGATCTCATTAAGAGCCTT GCCATGTTGAAGGACGGCTTCCGCACTGACAGGGGGTCTCACTCAGACCTGTGGTCCAGCAGC AGCTCTCTGGAGAGTTGAGTTTCCCGCTACCGAAACAGTACCTGGATGTGAGCTCCGACAGCA GACATCTCGGGAAGCTTCGGCATCAACAGCAACAACAGTTCGGCAGAGAAGGTAGATTGCGC CTTCGATATGAAGAGGCTAAGAGAAGGATCGCCAACCTGAAGTACAGAGTGGCCAAAGCTTGAC AGTGAGGCTGGCCTGGGGTGTCTGGACTCAGAGAGGGACCGGCTGATCCTTATCAACGAGAAG GAGGAGCTGCTGAAGGAGATGCGCTTCATCAGCCCCCGCAAGTGGACCCAGGGGGAGGTGGAG CAGCTGGAGATGGCCCCGAAGCGGCTGGAAAAGGACCTGCAGGCCAGCCCCGACACCCAGAGC AAGGCGCTGACGGAGAGGTTAAAGTTAAACAGTAAGAGGAACAGCAGTGTGAGAGAACTGGAG GAAGCCACCCGGCAGGTGGCAACTCTGCACTCCCAGCTGAAAAGTCTCTCAAGCAGCATGCA TCCCTGTCTCTCAGGCAGCAGCCCCGATCCCTCAGCTCCAGCCGGGCTCCCTGGTTGCATCC AGCCTGGACTCTCTCACTTCAGCCAGCTTCACTGACCTCTACTATGACCCCTTTGAGCAGCTG GACTCAGAGCTGCAGAGCAAGGTGGAGTTCCTGCTCCTGGAGGGGCCACCGGCTTCCGGCCC TCAGGCTGCATCACCACATCCAGCAGGATGAGGTGGCCACAGCCAGAGGCAGGAGGAGGT GGCCGCTGCAGGCTCTGCGTTCCCTGTCTGGCACCCCAAAGTCCATGACCTCCCTATCCCCA CGTTTCCTCTCTCTCTCCCCCTCCCCACCCTGTTCCCTCTCATGGCTGACCCCTCTTGCT GGTGATGCTTCTCTCAACTCTTGGAGTTTGAAGACCCGGAGCTGAGTGCCACTCTTTGTGAA CTGAGCCTTGTTAACAGCGCCAGGAAAGATACCGGCTGGAGGAACCAAGGACGAGGGGCAAG CAGCTGGGCCAAGCTGTGAATACGGCCAGGGGTGGGCTGAAAGTGGGCTGTGTCTCAGCC GCCGATCGGACGAGTCAGTGGCTGGAGACAGTGGTGTGTACGAGGCTTCCGTGCAGAGACTG GGTGCTTCAGAACTGCTGCATTGACAGTGACGAATCGGAAGCAGTGGGTGCGACCCGAATT CAGATTGCCCTGAAGTATGATGAGAAGAATAAGCAATTTGCAATATTAATCATCCAGCTGAGT AACCTTTCTGCTCTGTCGACGACAAGACAGAAAGTGAATATCCGCTGGCTGTCTTCTCT TGCTCTGAAAGCACAACACTGCCTGTTCCGGACCCGGCTCTGGAACGCTCAGACACTCTAGT TTCAATGAGGTGTTCTGGGTATCCATGTCTATCCAGCCCTTACCAGAAGACCTTAAGAGTC GATGTCTGTACCAACGACAGGAGCCATCTGGAAGAGTGCCTGGGAGGCGCCAGATCAGCCTG GCGGAGGCTGCGCGCTGTGGGAGAGGTCGACTCGCTGGTACAACCTTCTCAGCTACAATACT TTGAAGAAGCAGAGCAGGAGGCTCAAGCCAGTGGGAGTTATGGCCCTGCTCAGGCGCTGCC AGCACGGACGCTGTGTCTGCTCTGTTGGAACAGACAGCAGTGGAGCTGGAGAAGAGGCAGGAG GGCAGGAGCAGCACACAGACTGGAAGACAGCTGGAGGTATGAGGAGACCAGTGAGAATGAG GCAGTAGCCGAGGAAGAGGAGGAGGAGGTGGAGGAGGAGGAGGAGGAAGAGGATGTTTTACC GAGAAAGCCTCACCTGATATGGATGGGTACCCAGCATTAAAGGTGGAACAAAGACCAACAC GAGACCCCGGCTACCTTCCCCACAGTGGTGGCAAGCTAAGGACCGGAGAGTGGGACCCCGTCC CAGGGGCCATTCTTCGAGGGAGCACCATCATCCGCTCTAAGACCTTCTCCCCAGGACCCAG AGCCAGTACGTGTGCGGCTGAATCGGAGTGATAGTGACAGCTCCACTCTGTCCAAAAGCCA CCTTTTGTTTGAACCTCCCTGGAGCGACGACGCTCCGGATGAAGCGGCTTCTCTCGGTCAAG TCGCTGCGCTCCGAGCTCTGATCCGTACCTCGCTGGACCTGGAAGTTAGACCTGCAGGCGACA AGAAGCTGGCAGCAGCACTGACCCAGGAGATCTCGGTGCTGAAGTGAAGTCAAGGAGCAGCTG GAACAAGCCAAGAGCCACGGGAGAGGAGCTGCCACAGTGGTTGCGTGAGGACGAGCGGTTTC CGCCTGCTGCTGAGGATGCTGGAGAAGCGGATGGACCGAGCGGAGCACAAGGGTGAGCTTCAG ACAGACAAGATGATGAGGCGAGCTGCCAAGGATGTGCACAGGCTCCGAGGCGCAGAGCTGAAG GAACCCCAAGATTGAGTCTTTTTCAGGGAGAAGATGGCATTTTCCCTCCGCTCGGATGAT ATCCAGCTCTCTCTGCAGATGACGCTCTAATCGCCAGAAAGTATTTCTCTTGTTCACATGAC</p>

	CAGGCTGTGAACATTGACTGTGGCTAAAGTTATTTATGTGGTGTATATGAAGGTACTGAGTC ACAAGTCCTCTAGTGTCTTTGTTGGTTGAAGATGAACCGACTTTTGTAGTTGGGTCTACTG TTGTTATTAAAAACAGAACAAAAACAAACACACACACACAAAAACAGAAACAAAAAAAC CAGCATTAATAATAAGATTGTATAGTTTGTATATTTAGGAGTGTATTTTGGGAAAGAAAA TTTAAATGAACATAAGCAGTATTGAGTTGCTGCTCTTCTTAAATCGTTTAGATTTTTTTTGG TTTGTACAGCTCCACCTTTTAGAGGTCTTACTGCAATAAGAAGTAATGCCTGGGGGACGGTAA TCCTAATAGGACGTCCTCGCACTGTACAGTACAGCTAATTTTCTAGTTAACATATTTTGT ACAATATTAAAAAATGCACAGAAACCATTGGGGGGGATTACAGAGGTGCATCCACGGATCTTC TTGAGCTGTGACGTGTTTTATGTGGCTGCCAACGTGGAGCGGGCAGTGTGATAGGCTGGGT GGGCTAAGCAGCCTAGTCTATGTGGGTGACAGGCCACGTGGTCTCAGATGCCAGTGAAGCC ACTAATGAGTGAAGGGAGGGCTGTGGGAACTCCATTCAGTTTATCTCCATCAATAAAGT GGCCTTCAAAAAG		
	ORF Start: ATG at 94		ORF Stop: TAA at 3430
	SEQ ID NO: 256	1112 aa	MW at 125157.1kD
NOV72a, CG139224-01 Protein Sequence	MPRPELPPEGWEEARDFDGKVVYIDHTNRTSWIDPRDRYTKPLTFADCSIDELPLGWEEAY DPQVGDYFIDHNTKTTQIEDPRVQWRREQEHLKDYLVVAQEALSAQKEIYQVQQRLELAQQ EYQQLHAVWEHKLGSQSVSVSGSSSSSKYDPEILKAEIATAKSRVNLKREMVHLQHELQFKE RGFQTLKKIKDKMSDAQGSYKLDEAQAVALRETKAIKAITCGEKEKQDLIKSLAMLKDGFRFD RGSHSDLWSSSSSLESSSFLPKQYLDVSSQTDISGSFGINSNNQLAEKVRRLRYEFKAKRI ANLKIQLAKLDSEAWPGVLDSEDRDLILINEKEELLKEMRFISPRKWTQGEQLEMARKRLE KDLQAARDTQSKALTERLKLNSKRNQLVRELEEATRQVATLHSQLKSLSSSMQSLSSGSSPGS LTSSRGLSVASSLDSSTASFTDLYDPFEQLDSELQSKVEFLLLEGATGFRPSGCITTHIED EVAKTQKAEGGRLQALRSLSGTPKSMSTLSRSLSSSPSPCSPLMADPLLAGDAFLNSLEF EDPELSATLCELSLGNQAERYRLEEPGTEGKQLGQAVNTAQCGCLKVACVSAVDESVAAGD SGVYEASVQRLGASEAAAFDSDESEAVGATRIQIALKYDEKNQFALILIQSLNLSALLQQQD QKVNIRVAVLPCSESTTCLFRTRPLDASDTLVFNEVFWVMSYPALHQKTLRVVDCTTDRSHL EECLGGAQISLAEVCRSGERSTRWYNLLSYKYLKQSRLEKPVGVMAPASGPASTDAVSALLE QTAVELEKREQERSSTQTLEDSDRYEETSENEAVAEVEVEVEVEVEEDVFTEKASPMMDGY PALKVDKETNTETPAPSPPTVVRPKDRRVGTSPQGPFLRGSTIIRSKTFSPGPQSQYVCLNRS DSDSSTLSKKPPFVRNSLERRSVRMKRPSSVKSLRSELRIRTSLELDLQATRTWHSQLTQE ISVLKELKEQLEQAKSHGKELPQWLREDFRLLRLMLEKRMRAEHKQELQTDKMMRAAAK DVHRLRGQSKPEPEVQSFREKMAFFTTRPMNIPALSADDV		
	SEQ ID NO: 257	3062 bp	
NOV72b, CG139224-02 DNA Sequence	GCGGCCGCCCGGGCTAAGAGCGGCCGCTGGAGCCGCTGAGCCCCGCTGCGGCCGGGAGCTG CATGGGGGAGCGCCGCGCAGCGCTTGGGAAGATGCCCGCGCCGAGCTGCCCTGCGCGAGGGC TGGGAGGAGCGCGCGACTTCGACGGCAAGGTCTACTACATAGACCACAGAACCGCACACC AGCTGGATCGACCCGCGGGACAGGTACACCAAACCGCTCACCTTTGCTGACTGCATTAGTGAT GAGTTGCCGCTAGGATGGGAAGAGGCATATGACCCACAGGTTGGAGATTACTTCATAGACCAC AACACCAAACCACTCAGATTGAGGATCCTCGAGTACAATGGCGGCGGGAGCAGGAACATATG CTGAAGGATTACTGGTGGTGGCCAGGAGGCTCTGAGTGCAAAAGGAGATCTACCGAGTG AAGCAGCAGCGCCTGGAGCTTGACAGCAGGAGTACCAGCAACTGCATGCCGCTCTGGGAGCAT AAGCTGGGCTCCCAGGTCAGCGTGGTCTCTGGTTCATCATCCAGCTCCAAGTATGACCTGAG ATCCTGAAAGCTGAAATTGCCACTGCAAAATCCCGGGTAAACAAGCTGAAGAGAGAGATGGTT CACCTCCAGCAGAGCTGCAGTTCAAAGAGCGTGGCTTTCAGACCTGAAGAAGATCGATAAG AAAATGTCTGATGCTCAGGGCAGCTACAACTGGATGAAGCTCAGGCTGTCTTGAGAGAAACA AAAGCCATCAAAAAGGCTATTACCTGTGGGAAAAGGAAAAGCAAGATCTCATTAAAGAGCCTT GCCATGTTGAAGGACGGCTTCGCACTGACAGGGGCTCTCACTCAGACCTGTGGTCCAGCAGC AGCTCTCTGGAGAGTTTCAGTTTCCCGCTACCGAAACAGTACCTGGATGTGAGCTCCCAGACA GACATCTCGGAAGCTTCGGCATCAACAGCAACAATCAGTTGGCAGAGAAGGTGAGATTGGCG CTTCGATATGAAGAGGCTAAGAGAAGGATCGCCAACCTGAAGATCCAGCTGGCCAAGCTTGAC AGTGAGGCTGGCCTGGGGTCTGGACTCAGAGAGGACCGGCTGATCCTTATCAACGAGAAG GAGGAGCTGCTGAAGGAGATGCGCTTCATCAGCCCCGCAAGTGGACCGACCCCTCTGGCT GGTGATGCCTTCTCACTCTTGGAGTTTGAAGACCCGAGCTGAGTGCCACTCTTTGTGAA CTGAGCCTTGGTAACAGCGCCAGGAAAGATACCGCTGGAGGAACCAAGGAGGCAAG CAGCTGGGCAAGCTGTGAATACGGCCAGGGGTGTGGCCTGAAAGTGGCCTGTGTCTCAGCC GCCGTATCGACGAGTCACTGGCTGGAGACAGTGGTGTGTACGAGGCTTCCTGTCAGAGACTG GGTGCTTCAGAAGCTGCTGCATTGACAGTGACGAATCGGAAGCAGTGGGTGCGACCCGAATT CAGATTGCCCTGAAGTATGATGAGAAGAATAAGCAATTGCAATATTAACTACAGCTGAGT AACCTTCTGCTCTGTTGACGAACAAGACCAGAAAGTGAATATCCGCTGGCTGTCTTCTCT TGCTCTGAAAGCACAACTGCTGTTCGGACCCCGCTCTGGACGCTCAGACACTTAGTG TTCAATGAGGTGTTCTGGGTATCCATGTCTATCCAGCCCTTACCAGAAGACCTTAAGAGTC GATGTCTGTACCACCGACAGGAGCCATCTGGAAGAGTGCCTGGGAGGCGCCAGATCAGCCTG GCGGAGGTCTGCCGCTCTGGGAGAGGTCGACTCGTGGTACAACCTTCTCAGCTACAAATAC TTGAAGAAGCAGAGCAGGAGCTCAAGCCAGTGGGAGTTATGGCCCTGCTCAGGGCTGCC		

	AGCACGGACGCTGTGTCTGCTCTGTTGGAACAGACAGCAGTGGAGCTGGAGAAGAGGCAGGAG GGCAGGAGCAGCACACAGACACTGGAAGACAGCTGGAGGTATGAGGAGACCAGTGAAGATGAG GCAGTAGCCGAGGAAGAGGAGGAGGAGGTGGAGGAGGAGGAGGAGAAGAGGATGTTTTACC GAGAAAGCCTCACTGATATGGATGGGTACCCAGCATTAAAGGTGGACAAGAGACCAACAGC GAGACCCCGGCCCATCCCCACAGTGGTGCGACCTAAGGACCGGAGAGTGGGCACCCCGTCC CAGGGGCCATTTCCTCGAGGGAGCACCATCATCCGCTTAAGACCTTCTCCACGAGACCCAG AGCCAGTACGTGTGCCGGCTGAATCGGAGTGATAGTGACAGCTCCACTCTGTCCAAAAGCCA CCTTTTGTTCGAAACTCCCTGGAGCGACGACGCTCCGGATGAAGCGGCCTTCCTCGGTCAAG TCGCTGCGCTCCGAGCGTCTGATCCGTACCTCGCTGGACCTGGAGTTAGACCTGCAGGCGACA AGAACCTGGCACAGCCAAGTACCCAGGAGATCTCGGTGCTGAAGGAGCTCAAGGAGCAGCTG GAACAAGCCAAGAGCCACGGGGAGAAGGAGCTGCCACAGTGGTTGCGTGAGGACGAGGTTC CGCTGCTGCTGAGGATGCTGGAGAAGCGGATGGACCGAGCGGAGCACAAGGGTGAGCTTCAG ACAGACAAGATGATGAGGGCAGCTGCCAAGGATGTGCACAGGCTCCGAGGCCAGAGCTGTAAG GAACCCCCAGAAGTTCAGTCTTTCAGGGAGAAGATGGCATTTCACCCGGCTTCGGATGAAT ATCCAGCTCTCTCTGTCAGATGACGCTCTAATGCCAGAAAAGTATTTCTTTGTTCCACTGAC CAGGCTGTGAACATTGACTGTGGCTAAAGTTATTTATGTGGTGTATATGAAGGTACTGAGTC ACAAGTCCTCTAGTGCTCTTGTGGTTTGAAGATGAACCGACTTTTATGTTGGGTCTCTACTG TTGTTATTAAAAACAGAACAAAAACAAAACACACACAC		
	ORF Start: ATG at 94		ORF Stop: TAA at 2863
	SEQ ID NO: 258	923 aa	MW at 104821.7kD
NOV72b, CG139224-02 Protein Sequence	MPRPELPLPEGWEEARDFDKVYYIDHTNRTTSWIDPRDRYTKPLTFADCISDELPLGWEEAY DPQVGDFYFIDHNTKTTQIEDPRVQWRREQEHMLKDYLVVAQEALSAQKEIYQVKQORLELAQQ EYQQLHAVWEHKLGSQSVSVSSSSSKYDPEILKAEIATAKSRVNLKREMVLHQLHELQFKE RGFQTLKKIDKKIDMSDAQGSYKLEAQAQLRETKAIKAITCGEKEKQDLIKSLAMLKDGFRTE RGSHSDLWSSSSSLESSSFPLPKQYLDVSSQTDISGSFGINSNNQLAEKVRRLRYEEAKRRI ANLKIQLAKLDSEAWPGVLDSEDRDLILINEKEELLKEMRFISPRKWDPLLAGDAFLNSLEF EDPELSATLCELSLNSAQERYRLEEPGTEGKQLGQAVNTAQGCGLKVCVSAVSDSESVAGD SGVYEASVQRLGASEAAAFDSDESAVGATRIQIALKYDEKNQFAILIIQLSNLSALLQQQD QKVNIRVAVLPCSESTTCLFRTRPLDASDTLVFNEVFVWSMSYPALHQKTLRVDTVCTDRSHL EECLGGAQISLAEVCRSGERSTRWYNLLSYKYLKKQSRCLKPVGVMA PASGPASTDAVSALLE QTAVELEKRQEGRSSTQTLED SWRYEETSENEA VAEEEEEEVEEEG EEDVFTEKASPDMDGY PALKVDKETNTETAPSPPTVVRPKDRRVGTSPQGFRLRGSTIIRSKTFSPPQSQYVCLNRNS DSDSSTLSKKPPFVRNSLERRSVRMKRPSSVKSLRSELRIRTSLDLELDLQATRTWHSQLTQE ISVLKELKEQLEQAKSHGEKELPQWLREDERFRLLLRMLEKRM DRAEHKGELOTKMMRAAAK DVHRLRGQSCKEPPEVQS FREKMAFFT RPRMNI PALSADDV		
	SEQ ID NO: 259	3698 bp	
NOV72c, CG139224-03 DNA Sequence	GCGGCCGCGCGGGCTAAGAGCGGCGGCTGGAGCCGCTGAGCCCCGCTGCGGCCGGGAGCTG CATGGGGGAGCGCCCGCAGCGCTTGGGAAGATGCCCGCGCGGAGCTGCCCTTGCGGAGGGC TGGGAGGAGGCGCGGCACTTCGACGGCAAGGTCTACTACATAGACCACAGCAACCGCACCCACC AGCTGGATCGACCGCGGGACAGGTACACCAAAACCGCTCACCTTTGCTGACTGCATTAGTAGT GAGTTGCCGCTAGGATGGGAAGAGGCATATGACCCACAGGTTGGAGATTACTTCATAGACCAC AACACCAAAACCACTCAGATTGAGGATCCTCGAGTACAATGGCGGCGGGAGCAGGAACATATG CTGAAGGATTACCTGGTGGTGGCCAGGAGGCTCTGAGTGCAAAAAGGAGATCTACCAGGTG AAGCAGCAGCGCCTGGAGCTTGACAGCAGGAGTACCAGCACTGCATGCCGTCTGGGAGCAT AAGCTGGGCTCCAGGTGAGCGTGGTCTCTGTTTCATCATCCAGCTCCAAGTATGACCTGAG ATCCTGAAAGCTGAAATTGCCACTGCAAAATCCCGGGTAAACAAGCTGAAGAGAGAGATGGTT CACCTCCAGCAGAGCTGCAGTTCAAAGAGCGTGGCTTTCAGACCCTGAAGAAGATCGATAAG AAAATGTCTGATGCTCAGGGCAGCTACAACTGGATGAAGCTCAGGCTGTCTTGAGAGAAACA AAAGCCATCAAAAAGGCTATTACCTGTGGGAAAAGGAAAAGCAAGATCTCATTAAGAGCCTT GCCATGTTGAAGGACGGCTTCGCACTGACAGGGGTCTCACTCAGACCTGTGGTCCAGCAGC AGCTCTCTGGAGAGTTCGAGTTTCCCGCTACCGAAACAGTACCTGGATGTGAGCTCCAGACA GACATCTCGGGAAGCTTCGGCATCAACAGCAACAATCAGTTGGCAGAGAAGGTGAGATTGCCG CTTCGATATGAAGAGGCTAAGAGAAGGATCGCCAACCTGAAGATCCAGCTGGCCAAGCTTGAC AGTGAGGCTGGCCTGGGTGCTGGGACTCAGAGAGGAGACCGGTGATCCTTATCAACGAGAAG GAGGAGCTGCTGAAGGAGATGCGCTTCATCAGCCCCGCAAGTGGACCCAGGGGAGGTGGAG CAGCTGGAGATGGCCCGGAAGCGCTGGAAAAGGACCTGCAGGCAGCCCGGGACACCCAGAGC AAGGCGCTGACGGAGAGGTTAAAGTTAAACAGTAAGAGGAACAGCTTGTGAGAGAACTGGAG GAAGCCACCCGCGAGGTGGCACTCTGCACTCCAGCTGAAAAGTCTCTCAAGCAGCATGACG TCCCTGTCTCAGGCAGCAGCCCGGATCCCTCAGCTCCAGCCGGGGCTCCCTGGTTGCATCC AGCCTGGACTCCTCCACTTCAGCCAGCTTCACTGACCTCTACTATGACCCCTTTGAGCAGCTG GACTCAGAGCTGCAGAGCAAGGTGGAGTTCCTGCTCCTGGAGGGGGCCACCGGCTTCGGGCC TCAGGCTGCATCACCACCATCCAGGAGGATGAGGTGGGCCAAGACCCAGAAGGCAGGGAGGT GGCCGCTGCAAGCTCTGCGTTCCCTGTCTGGCACCCCAAAGTCCATGACCTCCCTATCCCA CGTTCCT		

	GGTGATGCCTTCCTCAACTCCTTGGAGTTTGAAGACCCGGAGCTGAGTGCCACTCTTTGTGAAG CTGAGCCTTGGTAACAGCGCCAGGAAAGATACCGGCTGGAGGAACAGGAACGGAGGGCAAG CAGCTGGGCCAAGCTGTGAATACGGCCAGGGGTGTGGCCTGAAAGTGGCCTGTGTCTCAGCC GCCGTATCGACGAGTCAGTGGCTGGAGACAGTGGTGTGTACGAGGCTTCCGTGCAGAGACTG GGTGCTTCAGAAGCTGCTGCATTTGACAGTGACGAATCGGAAGCAGTGGGTGCGACCCGAATT CAGATTGCCCTGAAGTATGATGAGAAGAATAAGCAATTGCAATATTATCATCCAGCTGAGT AACCTTTCTGCTCTGTTGCAGCAACAAGACCAGAAAGTGAATATCCGCGTGGCTGTCTCTCT TGCTCTGAAAGCACAACTGCCTGTTCCGGACCCGGCCTCTGGACGCCTCAGACACTCTAGTG TTCAATGAGGTGTTCTGGGTATCCATGTCCTATCCAGCCCTTACCAGAAGACCTTAAGAGTC GATGCTGTGTACCACCGACAGGAGCCATCTGGAAGAGTGCCTGGGAGGCGCCAGATCAGCCTG GCGGAGGTCTGCCGGTCTGGGGAGAGGTGCAGTCTGCTGGTACAACCTTCTCAGTACAAATAC TTGAAGAAACAGAGCAGGGAGCTCAAGCCAGTGGGAGTCATGGCCCTGCCTCAGGGCCTGCC CGGATGAAGCGGCCTTCTCGGTTAAGTCGCTGCGCTCCGAGCGTCTGATCCGTACCTCGCTG GACCTGGAGTTAGACCTGCAGGCGACAAGAACTGGCACAGCCAATTGACCCAGGAGATCTCG GTGCTGAAGGAGCTCAAGGAGCAGCTGGAACAAGCCAAGAGCCACGGGGAGAAGGAGCTGCCA CAGTGGTTGCGTGAGGACGAGCGTTTCCGCTGCTGCTGAGGATGCTGGAGAAGCGGCAGATG GACCGAGCGGAGCACAAAGGTGAGCTTCAGACAGACAAGATGATGAGGGCAGCTGCCAAGGAT GTGCACCAGCTCCGAGGCCAGAGCTGTAAGGAACCCCAAGAAGTTCAGTCTTTCAGGAGAAG ATGGCATTTTTTACCCGGCCTCGGATGAATATCCAGCTCTCTCTGCAGATGACGTCTAATCG CCAGAAAAGTATTTCTTTGTTCCACTGACCAAGGCTGTGAACATTGACTGTGGCTAAAGTTAT TTATGTGGTGTTATATGAAGGTACTGAGTCACAAGTCTCTAGTGCTCTTGTGTGTTTGAAGA TGAACCGACTTTTGTAGTTTGGGTCCTACTGTTGTTATTAATAACAGAACAAAAACAAACACA CACACACAAAAACAGAAACAAAAAAACCAGCATTAAAAATAAGATTGTATAGTTTGTAT TATTTAGGAGTGTATTTTGGGAAAGAAAATTTAAATGAAGTAAAGCAGTATTGAGTTGCTGC TCTCTTAAATCGTTAGATTTTTTTTGGTTTGTACAGCTCCACCTTTTAGAGGTCTTACTG CAATAAGAAGTAATGCCTGGGGACGGTAATCCTAATAGGACGTCCCGCACTTGTACACAGTAC AGCTAATTTTCTCTAGTTAACATATTTGTACAATATTAATAAATGCACAGAAACCATTGGG GGGGATTTCAGAGGTGCATCCACGGATCTTCTTGAGCTGTGACGTGTTTTATGTGGCTGCCCA ACGTGGAGCGGGCAGTGTGATAGGCTGGGTGGGCTAAGCAGCCTAGTCTATGTGGGTGACAGG CCACGCTGGTCTCAGATGCCAGTGAAGCCACTAACATGAGTGAGGGGAGGGCTGTGGGGAAC TCCATTAGTTTTATCTCCATCAATAAAGTGGCCTTTCAAAAAG		
	ORF Start: ATG at 94	ORF Stop: TAA at 2956	
	SEQ ID NO: 260	954 aa	MW at 107489.2kD
NOV72c, CG139224-03 Protein Sequence	MPRPELPPEGWEEARDFDGKVVYIDHTNRTTSWIDPRDRYTKPLTFADCSIDELPLGWEEAY DPQVGDFIDHNTKTQIEDPRVQWRREQEHMLKDYLVAQEALSAQKEIYQVKQORLELAQQ EYQQLHAVWEHKLGSQSVSVSGSSSSSKYDPEILKAEIATAKSRVNKLKREMVHLQHELQFKE RGFQTLKKIDKKMSDAQGSYKLDEAQAFLRETKAIKKAITCGEKEKQDLIKSLAMLKDGFRD RGSHTDLWSSSSSLESSSFLPKQYLDVSSQDISGSFGINSNNQLAEKVRRLRYEEAKRRI ANLKIQLAKLDSEAWPGVLDSEDRLLINEKEELLKEMRFISPRKWTQGEVEQLEMARKRLE KDLQAARDTQSKALTERLKLNSKRNLVRELEEATRQVATLHSQLKSLSSMSQLSSSGSPGS LTSSRGSVLASSLDSSTASFTDLYDPFEQLDSELQSKVEFLLLEGATGFRPSCGITTIED EVAKTQKAEGGRLQALRSLSGTPKSMSTLSPRSSLSPSPPCSPLMADPLLAGDAFLNSLEF EDPELSATLCELSLNSAQERYRLEEPGTEGKQLGQAVNTAQCGGLKVACVSAVSDSVAGD SGVYEASVQRLGASEAAAFDSDESEAVGATRIQIALKYDEKNKQFAILLIQLNLSALLQQD QKVNIRVAVLPCSESTTCLFRTRPLDASDTLVFNEVFVWSMSYPALHQKTLRVDVCTTDRSHL EECLGGAQISLAEVCRSGERSTRWYNLLSYKYLKKQSRELKPVGVMAPASGPARMKRPPSVKS LRSERLIRTSLDLELDLQATRTWHSQLTQEISVLKELKEQLEQAKSHGEKELPQWLREDFR LLLRMLEKQMDRAEHKGELQTDKMMRAAAKDVHQLRGQSCKEPPEVQSFREKMAFFTRPRMN IPALSADDV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 72B.

Table 72B. Comparison of NOV72a against NOV72b and NOV72c.		
Protein Sequence	NOV72a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV72b	553..1112 364..923	483/560 (86%) 483/560 (86%)

NOV72c	1..870	678/871 (77%)
	1..871	691/871 (78%)

Further analysis of the NOV72a protein yielded the following properties shown in Table 72C.

Table 72C. Protein Sequence Properties NOV72a	
PSort analysis:	0.7600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV72a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 72D.

Table 72D. Geneseq Results for NOV72a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV72a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU74354	Human cytoskeleton-associated protein (CYSKP) #25 - Homo sapiens, 912 aa. [WO200185942-A2, 15-NOV-2001]	202..1112 1..912	910/912 (99%) 910/912 (99%)	0.0
ABB11742	Human KIAA0869 protein homologue, SEQ ID NO:2112 - Homo sapiens, 894 aa. [WO200157188-A2, 09-AUG-2001]	225..1112 1..894	887/894 (99%) 887/894 (99%)	0.0
AAB93267	Human protein sequence SEQ ID NO:12300 - Homo sapiens, 379 aa. [EP1074617-A2, 07-FEB-2001]	734..1112 1..379	379/379 (100%) 379/379 (100%)	0.0
AAB42194	Human ORFX ORF1958 polypeptide sequence SEQ ID NO:3916 - Homo sapiens, 342 aa. [WO200058473-A2, 05-OCT-2000]	777..1112 1..342	335/343 (97%) 335/343 (97%)	0.0
AAB43089	Human ORFX ORF2853 polypeptide sequence SEQ ID NO:5706 - Homo sapiens, 202 aa. [WO200058473-A2, 05-OCT-2000]	1..170 23..192	167/170 (98%) 168/170 (98%)	7e-96

10

In a BLAST search of public sequence databases, the NOV72a protein was found to have homology to the proteins shown in the BLASTP data in Table 72E.

Table 72E. Public BLASTP Results for NOV72a				
Protein Accession Number	Protein/Organism/Length	NOV72a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O94946	KIAA0869 protein - Homo sapiens (Human), 888 aa (fragment).	225..1112 1..888	888/888 (100%) 888/888 (100%)	0.0
Q922W3	Unknown (Protein for IMAGE:3963643) - Mus musculus (Mouse), 967 aa (fragment).	145..1112 8..967	878/970 (90%) 909/970 (93%)	0.0
Q8VD17	Hypothetical 90.4 kDa protein - Mus musculus (Mouse), 812 aa (fragment).	293..1112 1..812	738/822 (89%) 764/822 (92%)	0.0
Q9BT29	Hypothetical 38.0 kDa protein - Homo sapiens (Human), 332 aa (fragment).	782..1112 1..332	331/332 (99%) 331/332 (99%)	0.0
Q8WVM4	Hypothetical 32.9 kDa protein - Homo sapiens (Human), 285 aa (fragment).	829..1112 1..285	284/285 (99%) 284/285 (99%)	e-159

- 5 PFam analysis predicts that the NOV72a protein contains the domains shown in the Table 72F.

Table 72F. Domain Analysis of NOV72a			
Pfam Domain	NOV72a Match Region	Identities/ Similarities for the Matched Region	Expect Value
WW	8..37	15/30 (50%) 27/30 (90%)	1.7e-11
WW	55..84	17/30 (57%) 25/30 (83%)	3.9e-07

Example 73.

- 10 The NOV73 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 73A.

Table 73A. NOV73 Sequence Analysis

	SEQ ID NO: 261	1455 bp	
NOV73a, CG140088-01 DNA Sequence	TCCTTGGCAGCCTTGGCAGCCACCTTGACACTCTCCTGTCTCCCCACCTCCACAGAGACAATG ACCATGTTTGAAAATGTCAACCGGCCCTGGCCAGACAGCTAAACCTCGAGGGGACCTGACA CCACTTGACAGCCTCATCGACTTCAAGCGCTTCCATCCCTTCTGCCTGGTCTGAGGAAGG AAGAGCACGCTCTTCTGGGGGGCCCGGTACGTCCGCACCGACTACACGCTGCTGGATGTGCTT GAGCCCGGCAGCTCACCTTCAGACCCAACAGACACTGGGAATTTGGCTTTAAGAATATGCTG GACACCCGAGTGGAGGGAGATGTGGATGTACCAAAGACGGTGAAGGTGAAGGGAACGGCAGGG CTCTCGCAGAACAGCACTCTGGAGGTCCAGACACTCAGTGTGGCTCCCAAGGCCCTGGAGACC GTGCAGGAGAGGAAGCTGGCAGCAGACCACCCATTCTGAAGGAGATGCAAGATCAAGGGGAG AACCTGTATGTGGTGTGGAGGTGGTGGAGACGGTGCAGGAGGTCACTGGAGCGAGCCGGC AAGGCAGAGGCCTGCTTCTCCCTCCCTTCTTCGCCCCATTGGGGCTACAGGGATCCATAAAT CACAAGGAGGCTGTAACCATCCCCAAGGGCTGCGTCTGCGCTTTCGAGTGAGACAGCTGATG GTCAAAGGCCAAAGATGAGTGGGATATTCACATATCTGCAATGATAACATGCAACCTTCCCT CCTGGAGAGGAAGGTGCCGATGGAGGGCTGTGTGTCCCATTTATGTCCCATAGGGGACGTA CACGAAGGCTTCAGGACACTAAAAGAAGAAGTTCAGAGAGAGACCAACAAGTGGAGAAGCTG AGCCGAGTAGGGCAAAGCTCCCTGCTCAGCTCCCTCAGCAAACTTCTAGGGAAGAAAAAGGAG CTACAAGACCTTGAGCTCGCACTTGAAGGGCTCTAGACAAGGGACATGAAGTGACCTGGAG GCACTCCCAAAAGATGTCTGCTATCAAAGGAGGCGGTGGGCGCCATCTCTATTTCTGTGGA GCCCTAACAGAGCTAAGTGAAGCCCAACAGAAGCTGCTGGTGAATCCATGGAGAAAAAGATC CTACCCGTGCAGCTAAAGCTGGTGGAGAGCAGATGGAACAGAACTTCTGCTGGATAAAGAG GGTGTTTTCCCTTGCAACCTGAGCTGCTCTCTCCCTTGGGACGAGGAGTGACCTCAGC GAGGCTCTAGTCCGGCTGAGTGGCTGGAAGTGACAGAGATCGGGCCCCCAATATATGTGGGAC CCAGACACCCTCCCTCGCTCTGTGCTCTTTATGCAGGCCTCTCTCTCTCAGCAGCTTACC AAGGCCTCTAATTTGCCTTTACGTCTGCTTCATGACTCCCTAATGCCTTCCCAACCTCGTG GTGCTG		
	ORF Start: ATG at 61		ORF Stop: TAA at 1396
	SEQ ID NO: 262	445 aa	MW at 49428.5kD
NOV73a, CG140088-01 Protein Sequence	MTMFENVTRALARQLNPRGDLTPLDSLIDFKRFPFLVLRKRKSTLFWGARYVRTDYTLDDV LEPGSSPSDPTDTGNFGFKMLDTRVEGDVDVPKTVKVKGTAGLSQNSTLEVQTLVAPKALE TVQERKLAADHPFLKEMQDQENLYVMEVETVQEVTLERAGKAEACFSLPFFAPLGLQGS I NHKEAVTIPKGCVLAFVRQLMVKGKDEWDIPHICNDNMQTFPPGEGARWRVCPIIIVIGD VHEGFRTLKEEVQRETQVEKLSRVQSSLLSSLSKLLGKKKELQDLELALEGALDKGHEVTL EALPKDVLSSKEAVGAILYFVGALTELSEAQQKLLVKSMEKKILPVQLKLVESTMEQNFLLDK EGVFPLQPELLSSSLGDEELTLTEALVGLSGLEVQRSGPQYMWDPDTLPRLCALYAGLSLLQQL TKAS		
	SEQ ID NO: 263	1386 bp	
NOV73b, CG140088-02 DNA Sequence	CTCCACAGAGACAATGACCATGTTTGAAAATGTCAACCGGCCCTGGCCAGACAGCTAAACCC TCGAGGGGACCTTGACACCCTTGACAGCCTCATCGACTTCAAGCGCTTCCATCCCTTCTGCCT GGTGCTGAGGAAGAGGAAGAGCAGCTCTTCTGGGGGGCCCGGTACGTCCGCACCGACTACAC GCTGCTGATGTGCTTGAGCCCGGAGCTCACCTTCAGACCCAACAGACACTGGGAATTTGG CTTAAGAATATGCTGGACACCCGAGTGGAGGGAGATGTGGATGTACCAAAGACGGTGAAGGT GAAGGGAACGGCAGGGCTCTCGCAGAACAGCACTCTGGAGGTCCAGACACTCAGTGTGGCTCC CAAGGCCCTGGAGACCTTGCAAGAGAGGAAGCTGGCAGCAGACCAACCCATTCTGAAGGAGAT GCAAGATCAAGGGGAGAACCTGTATGTGGTGTGGAGGTGGTGGAGACGGTGCAGGAGGTAC ACTGGAGCGAGCCGGCAAGGCAGAGGCCTGCTTCTCCCTCCCTTCTTCGCCCCATTGGGGCT ACAGGGATCCATAAATCACAAGGAGGCTGTAACCATCCCCAAGGGCTGCGTCTGCGCTTTTCG AGTGAGACAGCTGATGGTCAAAGGCAAAGATGAGTGGGATATTCACATATCTGCAATGATAA CATGCAAACTTCCCTCCTGGAGAAAAGTCAAGAGAGGAGAAGGTATCCTTATCCAGGCATC TGATGTTGGGGACGTACACGAAGGCTTCAGGACACTAAAAGAAGAAGTTCAGAGAGAGACCA ACAAGTGGAGAAGCTGAGCCGAGTAGGGCAAAGCTCCCTGCTCAGCTCCCTCAGCAAACTTCT AGGGAAGAAAAAGGAGCTACAAGACCTTGAGCTCGCACTGAAGGGCTCTAGACAAGGGACA TGAAGTGACCTTGAGGGCACTCCCAAAAGATGTCTGCTATCAAAGGAGGCGGTGGGGCCCAT CCTCTATTTCTGTGGAGCCCTAACAGAGCTAAGTGAAGCCCAACAGAAGCTGCTGGTGAATC CATGGAGAAAAAGATCCTACCCGTGCAGCTAAAGCTGGTGGAGAGCAGATGGAACAGAATT CCTGCTGGATAAAGAGGGTGTTCCTCCCTGCAACCTGAGCTGCTCTCTCTCTGGGGACGA GGAGCTGACCTTACGGAGGCTCTAGTGGGCTGAGTGGCTGGAAGTGACAGATCGGGCCC CCAATATATGTGGGACCCAGACACCCTCCCTCGCTCTGTGCTCTTTATGCAGGCCTCTCTCT CCTTCAGCAGCTTACCAAGGCCTCTAATTTGCCTTTACGTCTGCTTCATGACTCCCTAATG		
	ORF Start: ATG at 14		ORF Stop: TAA at 1349
	SEQ ID NO: 264	445 aa	MW at 49377.4kD

NOV73b, CG140088-02 Protein Sequence	MTMFENVTRALARQLNPRGDLTPDLSLIDFKRFHPFCLVLRKRKSTLFWGARYVRTDYTLDDV LEPGSSPSDPTDTGNFGFKNMLDTRVEGDVDVPKTVKVGKTAGLSQNSTLEVQTLSPVAPKALE TLQKRKLAADHPFLKEMQDQGENLYVMEVVETVQEVTLERAGKAEACFSLPFFAPLGQSGI NHKEAVTIPKGCVLAFRVRQLMVKGKDEWDIPIHCNDNMQTFPPGEKSGEEKVILIQASDVGD VHEGFRTLKEEVQRETQQVEKLSRVGQSSLLSSLSKLLGKKKELQDLELALEGALDKGHEVT EALPKDVLLSKEAVGAILYFVGALTELSEAQQKLLVKSMKKILPVQLKLVSTMEQNFLLDK EGVFPLQPELLSSSLGDEELTLTEALVGLSGLEVQRSGPQYMWDPTLPRLCALYAGLSLLQQL TKAS
--	---

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 73B.

Table 73B. Comparison of NOV73a against NOV73b.		
Protein Sequence	NOV73a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV73b	1..445	398/445 (89%)
	1..445	404/445 (90%)

5

Further analysis of the NOV73a protein yielded the following properties shown in Table 73C.

Table 73C. Protein Sequence Properties NOV73a	
PSort analysis:	0.3600 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.3000 probability located in nucleus; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV73a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 73D.

Table 73D. Geneseq Results for NOV73a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV73a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB93904	Human protein sequence SEQ ID NO:13862 - Homo sapiens, 484 aa. [EPI074617-A2, 07-FEB-2001]	4..443 5..481	157/480 (32%) 242/480 (49%)	3e-50
ABB90142	Human polypeptide SEQ ID NO 2518 -	4..433 5..469	149/470 (31%) 233/470 (48%)	6e-45

	[WO200190304-A2, 29-NOV-2001]			
AAB66866	Human peptidyl-prolyl isomerase-2 - Homo sapiens, 443 aa. [US6171843-B1, 09-JAN-2001]	4..373 5..410	128/410 (31%) 202/410 (49%)	9e-34
ABB97522	Novel human protein SEQ ID NO: 790 - Homo sapiens, 403 aa. [WO200222660-A2, 21-MAR-2002]	2..443 3..396	134/462 (29%) 219/462 (47%)	1e-30
ABB72295	Murine protein isolated from skin cells SEQ ID NO: 507 - Mus sp, 244 aa. [WO200190357-A1, 29-NOV-2001]	247..445 23..232	82/212 (38%) 118/212 (54%)	3e-23

In a BLAST search of public sequence databases, the NOV73a protein was found to have homology to the proteins shown in the BLASTP data in Table 73E.

Table 73E. Public BLASTP Results for NOV73a				
Protein Accession Number	Protein/Organism/Length	NOV73a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
BAC04790	CDNA FLJ39120 fis, clone NTONG2006646, highly similar to Mus musculus Gd gasdermin - Homo sapiens (Human), 445 aa.	1..445 1..445	428/445 (96%) 433/445 (97%)	0.0
Q96QA5	Gastric cancer-related protein FKSG9 - Homo sapiens (Human), 446 aa.	1..445 1..446	421/446 (94%) 428/446 (95%)	0.0
Q9EST1	Gasdermin - Mus musculus (Mouse), 446 aa.	1..445 1..446	382/446 (85%) 413/446 (91%)	0.0
Q9D810	2200001G21Rik protein - Mus musculus (Mouse), 276 aa.	186..445 19..276	191/260 (73%) 220/260 (84%)	e-101
Q9D8T2	1810036L03Rik protein - Mus musculus (Mouse), 487 aa.	4..441 5..482	152/481 (31%) 243/481 (49%)	1e-47

5

PFam analysis predicts that the NOV73a protein contains the domains shown in the Table 73F.

Table 73F. Domain Analysis of NOV73a			
Pfam Domain	NOV73a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 74.

The NOV74 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 74A.

Table 74A. NOV74 Sequence Analysis			
	SEQ ID NO: 265	11124 bp	
NOV74a, CG140170-01 DNA Sequence	GAAATTTTAATTTATTATTCCTCCCTTTTCTTCTGTCATCTATAGGATAATATTGTAAATAG CAATGAAACCAATAATCATTAAATAAATATCAAGGAAAATCCAAGCAAAGCTTCTTTTGT TGGACTAGTGGTGTGGTGTGGAGACAGTCTCTGAATGTGAACAGGAAAGCACCATCAGCA AAACACGATCACTCTTAGGGAGACAGCTGGGGGAATCTGACTCTGGCTTCTGCTTTGTTTT AAGGGATTAACTTCCCTGTCAAGTCCAAGAAGACTTGCATGAGAAGATTACCTGATGGACT TAATCTAAGATTAGCTTTTTCATCAAGATGGAAGATCTTTAGGAGCAGAAAAGGGGAG TGCTAACTGGGGGAGCGAGAAGGGAGACGAGCAAAAGAAACAAATCTTGCCACGTGGCTCTG TTTGTGTCAGCAAGAGGATTTAAGACTCACCCAGGGCAAACTGGGACCACTGTAAAGACGCT GGAACATTCTGCCTCTTGAGTGAAGGGCCCTCTTCTAGCCTCTATGGCACTGAGGGGTGCG CCGGCTGGTGGAGGAGTAGTCCGATGGAGCCCTGCGTTCCCGGGGACACAGGGCCAAGCTT GAGGTGGAAGTTTCTGGTTCTGAAACAACAAGGAGAGAGTCTGTTTTCTTCTCTAAAATTG GACTCTTGTCTGCACAACTCTGGTCTGTTTGCACGGTTTGTGTGCCTTTTTTCCCTTTAT GCAATCTTTTTCAGCTTTAGCAGCAGAAATTGTCTAGTTCAAGAAACATGCTAGAGGGTGGC TTCAGAAGGAAGATGATCCTGTGTATTCTGTCTGTCATCCGAACTTTTGAAGAGAAAATTCT GAGCTAGAGGGATTCTTAAAGCCTTAAGTTACTTGAATCTATGTATTGTCAACCCCTTTGTCT CTGGAATCATATTACACTAACTGGAATCTCAGGCTGAATGAGAATAACCAAGTGGAGTAAAA AGAAGAAAACCGTTTCTTGATCACCACTTAATTAACGATGCTCTTCTCCAAAGGATCAGCAC GTTCTTCCCTCTGAGAACTTGAAAATACAAATGGACCCATGTTTTTTAAGCATTACCTTTTC TTAGAAGACTGCCATCATCTTTATAGAGGAATTTTTTCACTATGCATTGAGTGATCTTTAT AAAATACTGACCTTCTAATTAGATTAGGTGAGTCTTAATTAAGGGGGAAAAAGCAACGCA AGCCAACCACAAAAACACATATACCAATGAAAGAAATTTGGTTAAATTTACAGCATTAAACAT TACTTTTTAAGTAAACAGTTCATTGAAGAAAGTATGTATGCAGCAGTGGAAACATGGCCCTGT GCTTGTGAGTGACTCCAACATCCTGTGCCTGTCTGGAAGGGCGTGTCCCCAAGAGTGAGAA GGAGAAGCCTGTGTGCAGGAGACGCTACTATGAGGAAGGCTGGCTGGCCACGGGCAACGGGCG AGGAGTGGTTGGGGTGACTTTCACTCTAGTCACTGTGCGAGGGACAGGAGTACTCCACAGAG GATAAATTTCAACCTCCGGGGCCACAATAGCGAGGTGTGCTGGTGGAGTGGAAATGAGCCCTA CCAGAACTGGCCACGTGCGATGCGGACGGAGGCATATTCGTGTGGATTCAGTACAGGGCAG GTGGTCTGTGGAGCTGGTCAACGACCGGGGCGCAGGTGAGTGATTTCAGCTGGAGCCATGA TGGAACCTCAAGCACTTATTTCTATCGAGATGGGTGTGCTGGTGGGTCTGTCAGTGGACA AAGACACTGGTCAATCCGAAATCACTTGGAAAGTCAAATTACGTGTGGCATATGGACTCCTGA CGACCAACAGGTGTGTTTGGCACGGCCGATGGGCAGGTGATGTGTCATGGATTGCCACGGCAG AATGCTGGCCACGTCTCTTGCACGAGTCAGACGGTGTCTCGGCATGTCTGGAACCTACCC GATCTTCTGGTGGAGGACAGCAGCGAGAGCGACAGGACTCAGATGACTACGCCCCCCCCA AGATGGTCCGGCAGCATATCCATCCAGTGCAGAACATCAAGCCTCTGCTACCCGTGAGCTT CACCTCGGGAGACATCAGCTTAATGAACAACTACGATGACTTGTCTCCACGGTCACTCCGCTC AGGGCTGAAAGAGGTGGTAGCCAGTGGTGACACAGGGGACTTGTGGCAGTCTGCTGGGAT GGAACGGCAGACCAGCTTGGTGAGCTTCCCAATGGTCCCCTCTGAAGAGTGCCATGGTCAA GTTCTACAATGTTTCGTGGGGAGCAGATCTTCACTGAGCACTCTCGTGCAGCGCCCCATCAT CTCCATCTGCTGGGGTACCAGGATTCGAGGCTGTTGATGGCATCAGGACCAGCCCTGTACGT GGTGGTGTGGAGCACCGGGTGTCCAGCCTGCAGCTGCTGTGCCAGCAGGCCATCGCCAGCAC CTTGGCTGAGGACAAGGACGTGAGCAAGCTGACTCTGCCCCCGGCTCTGTCTCTACCTCTC CACTGCCTTCATCCCACCATCAAGCCCCCAATTCAGATCCGAACAACATGAGAGACTTTGT CAGCTACCCATCAGCCGGCAACGAGCGGTGCACTGCACCATGAAGCGCACAGAGGACGACCC GGAGGTGGGCGGCCGTGCTACACGCTCTACCTGGAGTACCTGGGCGGGCTGTGTGCCATCTC CAAAGGGCGGCGCATCAGCAAGCTGCGGGCAGAGTTCGTATCATGAGACCCCGGACAGATAG CAAACCAGATGAAATCTATGGGAACAGCTTGATTCTACTGTGATCGACAGCTGCAACTGCTC AGACTCCAGTGACATTGAGCTGAGTGATGACTGGGCTGCCAAGAAATCTCCAAAATCTCCAG AGCTAGCAAATCACCCAACTCCCAAGGATCAGCATGAGGCCCGCAAGTCAACCAAGCTGCC CCGGGCTGCTCAGGAGCTTCCCGGTCCCACGGTGGCCCTGCGCAAGCCCTCTGTGGGCTC GCCAGCTGACTCGGAGAGAGTTTCTTTTGAAGACATCACTACCCCACTATCTTGCTCA GGTCACGTCTAATATCTGGGAACCAAATTAAGATTGTGGGCTTGGCTGCTTCTGCCAAC CAACCTCGGTGAGTAATCTATAAAACAGCCTCTGCATCTCAGCCCGCGCAGATGACCAT TTATCTCCAGAAGTTCGAAAAATTCATGGACTATATTAATTTACCTGTCTTCAACCCAAA TGTTTTAGTGAAGATGAAGATGATTTACCAGTGACAGGAGCATCTGGTGTCCCTGAGAACAG CCCACCTTGTACCGTGAACATCCCTATTGCACCGATCCACAGCTCGGCTCAGGCTATGTCCCC		

374

	<p>AGTATTTTCAAATTTAAACGTTTTTAAATCCCCTGTTTAGTTAAAAAATGGAAAAGAAACC GACCCATTTTCCCAGATCAAGATGACATGACATCACTCCCAATTCTCTCCAAACCCGAGA GAAATACTGACGAAGTTTCTGATGTGGCAAAGGATATTTCCCATCTAATACCAAGTTTCTCAT TTATATTTAACGTATTGGACCTGATATTTTAGTGGGTGCATTCTTCCAGAAAGAAATTCAGCA ATGTATATCAGAATTAAATCTTTATATGAGTTTATGTAGCTTGATATGGTGTTCAGTGCTTA TTGGTTGTGCAATAATGGTTATAGCCTGTTAGATAATCTAAATGCAATTTCCCTGTTTTGTCTG TTTAGGAGATAATTATTTATCTTGCTTTTCATAGTGTCTTAGGAATTATTTTGTGTGTACGT TTTGGTGAGTTATACCCATTTATTTATTTAGAAAAATAGTATCTTTGTTAACGACTTACATG GTCACAGTATATTTTGTGCAAGAAATAAGAGGATATGATAGAAGGTTTTTTTTTTTTTTTT TTTTTTTTTTTGAGACGGAGTCCCACTCTTGTGCCCCAAGTAGAGTGTAGTGGCACAATCTCG GCTCCCCACAACCTCTGACTCCAGGGTTCAGGTGATTATTTTGCCTCAGCCTCCCAAGCAGCT GGGATTATAGACACCGCCCAACACGCCAGGCTAATGTTTTGTATTTTTAATAGAGATGGGGT TTTGCCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCGCCCCCTCGGCCT CCCAAAGTCTGGGATTACAGACGTGAGCCACCACTCCCGGCCCATAGAAGGTTTTTGTCTGG ATAATTTGTAACTTTCTAATGGGAAAAAATTCCTAATCACTTAAAAATTTTTTTTGTG ATTTTGTGTATTGATTATATACAAAGGAGACTTTTTTTTGAGATAACACTCAAATAGTATT CTCTTCTTTGAAAAATTTATTTTATCTGAAAAATAACAGTTGATCTGAAATAAAAAGGGGAG ACCTATTAGAATGAGAGTAGCCAAGGAAAGAGTTACTAGGTAATAAGCTTCACTTTTTGTGTT CTAATTGTTTTTGAGATATAAAGACCCCTGAAAAAGCCCATTTAGAACCTGTTAATAAGAGC AAATATAGGGGAAAAATCTTTGAAATGAAAGCTACAAATACATGTGAGAAGAAAAAATGGATT TTTTAGCAAATAAATAACTAAGCTTCTAAATGCCTAGCTCCCTCCCAAGGCGCTTTCCC CCGATGGAGGCACAGGCTTCTGTCTCGGATGTTTGGCGCACGTGAGTTGTATGAGTTGTAC CGGAGTGACCCCGCAGCCACTGCCACCTCCCTCTACCCAGGGGCTGAAAGAGGGGCTG CCCTCTGCGCCAAGGCAGACACAGCTGCGGGCTGTGCGGTCTAGTAGTGTGACGTTTCAG TTAATAGTGGTGGTCTTATTTCAACTATGCTTTCATTCAGTCAGTCTCTGTTGACTAAATAC GACGAAAAATTCATACTTTATGCAGGAGATTCTAAAAATTTAATGTTTATTAATAGTTTATGA ATATCAAGATACCTCATTGAATCCCTAAATTTAAAGCAGTCCAGTAAAGGTTAACTGTATA AAGAATCTATGACTTTTTGAGGGAAGTGTGATATATTAACAAATATAACCAATTTCTAAATTG TTTTAGCTCTAACCTCATCAAACCAAGGCACAGATTGTGTACAATATACCCATTGAATGTA TATCCTGAGAAAAATTTGGGCCAAAGAAGCAGGAAAAATCTCAAAGCTAATGGCAGCATAAA TCAAAGAATTTACAGGCTAGTGTTTTTATCCATAGCCATTGCTCCCTGTCAAGTGTCTCAC AAGGACATGGAAGAATGTGTATGTTCATCTGTAAATCATAGCAAAAAGCTGCAAAACCCGAG GGTCAAGCCTGCTCTGCCACAGGGTTGGATGGTGACCTTGGGCAAGTCCCTGGGGCTGGCTAG GCCTCCACTTGTCCATCTGTGAAATGAAAGGATCAGCCTGGACAGCCCTCTAACTCCCTTAC AGCTCTCAGCCTAAGAGCGCAGCACTGAACAGCCTCATCATTTCCACTTTTCATGGGAAATATA TTTCACACCATTGCCTTTGTGTAGAGAAATATTTCTTTCTGTGTTAATGAGCTATGTACTG AATATAAACCAAGTGCATTAAAGTAATATCTTTTGTGCACCTCTAAATGTGTTTGAATTTGTG TTTGTCTCATAGAAATATACAAAGTACTGATTCTAGGTAAGAAGGAGTCTCCACGGGTGTGC CCTGCTCAGCTGGATGTCCATGAGAACAGCCATGAAATAAGTCACTACTGTCCCCAAACCA CAGGAATATATACCTAGGTCACCTCAAATTCCTGAGTGTGCTCTGCCATGTTACACGGTCTTC AAATTGAAAGGTTTCTTGAAAGGAAAGTTTGGCCAGCACTGGAGAAGTGTGTTGCAATGCTT TCGCTGTGTGCTGTATCATTGGCCAAGTCAATGGTGTGAAGCAAAGTTAGGTGGAGACAAAA ATGTGTCCAAATGTCGTTGAGTTCTTGGGATTTCTGTAATAGCACACAACCTCAGAACTCTT CAGCATTGTGTGATTCTTACCTCTGGCTGATAAACTCTAATGGGTGTGGCTTACTTTGT TTCCATTTCTTTGGCTTTGTGCAATTTTGTGTAACCTTACTTGTACCTATATTTCTGTTT ACAGTTCTTTTAAAGGGGAGGGGTAGGGTTCTAAGATCTGTTGTTTATTTAGATAGAAAAAT TTTTCGTGTGTAGAAAAGCATGGGTATGCGTTTGAAGTGAAGAAAGACACTGTATTATTTACC AAAGGGGTATTGTTTTGCATTTGTTTATAAATGCATTATTTTGGTACTGTAATTTGGACAT AATTTCTGAGTTTATTACTACTGGCATTTCCTTTTTCCCTTTTTTTTTTTTTTAAACCGTAAGT GCACGATGCAGGTGCATAGGCCCCAGACCAAACTAGACCACCAGCATGTTCTGTCAGACCT CGGCAGTGGCGTGCATGCTTGTGCACCTCAGTTCCTCCAGTGTGGTGTGTTGTTTGTAAAA TTCAGCATCCTGCTGGTTTACTTTCCAAGCAAGATCTGTGCGACTCCCAATGCGTTTAA TGAGCTCATCCTATTTCCTTTCTTCTTACGTATTTTGTGATTAGATTGTGCAGGAGATAT TCTAGAAGGCATTAATGGTTTGCAATTTAAAACGATGTGGTTGTCCAAGTTATTTCTGCTCTT TATTACTGAGACGGATTAACTCTCTTATTTTTTTCTTGATGATTGAAGTTGAAGAGTTGTC CAGCTATTGCTTAATAAAATTTTGCAGATCAAAAAA</p>
	<p>ORF Start: ATG at 1358</p>
	<p>ORF Stop: TGA at 5987</p>
	<p>SEQ ID NO: 266</p>
	<p>1543 aa</p>
	<p>MW at 168953.6kD</p>
<p>NOV74a, CG140170-01 Protein Sequence</p>	<p>MYAAVEHGPVLCSDSNILCLSWKGRVPKSEKEKPVCRRRYYEEGLATGNRGVGVFTSSH CRRDRSTPQRINFNLRGHNSEVVLVRWNEPYQKLATCDADGGIFVWIQYEGRWSVELVNDRA QVSDFTWSDGTQALISYRDGFLVGVSVGQRHWSSEINLESQITCGIWTDDQQLVFGTADG QVIVMDCHGRMLAHVLLHESDGLVGMWNYPIFLVEDSSESDDTSDDDYAPQDGPAPYPIPVQ NIKPLLTVSFTSGDISLMNMYDLSPTVIRSLKEVVAQWCTQGDLLAVAGMERQTQLGELP GPLLKSA MVKFFYNVRGEHIFLDTLVQRPIISICWHRDRSRLLMASGPALYVVRVHRVSSLO LLCQQAIASTLREDKDVSKLTLPPRLCSYLSAFIPTIKPPIPDNNMRDFVSYP SAGNERLH</p>

CTMKRTEDDPEVGGPCYTLYLEYLGGLVPIILKGRRISKLRPEFVIMDPRTDSKPDEIYGNLSL STVIDSCNCSDDSDIELSDWAAKKSPKISRASKSPKLPRISIEARKSPKLPRAAQELSRSPR LPLRKPSVGSPSLTRREFPFEDITHPTYLAQVTSNIWGTKFKIVGLAFLPTNLGAVIYKTSL LHLQPRQMTIYLPEVRKISMDYINLPVFNPNVFEDEDDLPTVGASGVPENSPPCTVNIPIAP IHSSAQAMSPTQSIGLVQSLLANQNVQLDVLTNQTTAVGTAEHAGDSATQYPVSNRYSNPGQV IFGSEVMGRIIQNPPLSLPPPPQGMQLSTVGHGDRDHEHLQSAKALRPTPQLAAEGDAVV FSAPQEVQVTKINPPPPYPGTIPAAPTTAAPPPLPPPQPPVDVCLKKGDFSLYPTSVHYQTP LGYERITTFDSSGNVEEVCPRTRMLCSQNTYTLPGPGSSATLRLTATEKKVPQPCSSATLNR LTVPRYSIPTGDPPEYPIASQLAQGRGAQRSDNSLIHATLRRNNREATLKMAQLADSPRAP LQPLAKSKGGPGGVVTLPARPPPALYTCSQCSGTGPSSQPGASLAHTASASPLASQSSYLL SPDSARDRTDYVNSAFTEDALSQHCQLEKPLRHPLPEAAVTLKRPPPYQWDPMLGEDVWV PQERTAQTSGPNPLKLSSLMLSGQHLDVSRLPFISPKSPASPTATFQTGYGMGVYPGYSYN PPLPGVQAPCSPKDALSPQFAQQEPVVLPPLYPPLSYCTLPMPYPGSSSTCSSLQLPVAL HPWSSYSACPPMQNPQGTLPKPHLVVEKPLVSPPPADLQSHLGTEVMVETADNFQEVLSLSE SPVQORTEKFGKKNRRLDSRAEESVQAITEGKVKKEARTLSDFNSLISSPHLGREKKVKVS QKDQLKSKKLNKTNEFQDSSESEPELFISGDELMNQSQGSRKGWKSRRSPRAAGELEEAKCRR ASEKEDGRLGSQGFVYVMANKQPLWNEATQVYQLDFGGRVTQESAKNFQIELEGRQVMQFGRI DGSAYILDFQYPFSAVQAFALANVTORLK
--

Further analysis of the NOV74a protein yielded the following properties shown in Table 74B.

Table 74B. Protein Sequence Properties NOV74a	
PSort analysis:	0.8800 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV74a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 74C.

Table 74C. Geneseq Results for NOV74a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV74a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB97398	Novel human protein SEQ ID NO: 666 - Homo sapiens, 678 aa. [WO200222660-A2, 21-MAR-2002]	1..671 1..671	668/671 (99%) 668/671 (99%)	0.0
ABB61656	Drosophila melanogaster polypeptide SEQ ID NO 11760 - Drosophila melanogaster, 1478 aa. [WO200171042-A2, 27-SEP-2001]	37..669 86..702	302/650 (46%) 393/650 (60%)	e-150
AAM92361	Human digestive system antigen SEQ	417..495 1..79	79/79 (100%) 79/79 (100%)	9e-42

	[WO200155314-A2, 02-AUG-2001]			
ABB10608	Human pancreatic cancer related polypeptide, SEQ ID NO: 257 - Homo sapiens, 99 aa. [WO200155206-A1, 02-AUG-2001]	417..495 1..79	79/79 (100%) 79/79 (100%)	9e-42
ABB58828	Drosophila melanogaster polypeptide SEQ ID NO 3276 - Drosophila melanogaster, 1205 aa. [WO200171042-A2, 27-SEP-2001]	20..370 5..338	102/366 (27%) 169/366 (45%)	4e-29

In a BLAST search of public sequence databases, the NOV74a protein was found to have homology to the proteins shown in the BLASTP data in Table 74D.

Table 74D. Public BLASTP Results for NOV74a

Protein Accession Number	Protein/Organism/Length	NOV74a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9NRJ4	Tubby superfamily protein - Homo sapiens (Human), 1544 aa.	1..1543 1..1544	1534/1544 (99%) 1534/1544 (99%)	0.0
Q9JIL5	Tubby superfamily protein (Tubby-like protein 4) - Mus musculus (Mouse), 1547 aa.	1..1543 1..1547	1455/1547 (94%) 1491/1547 (96%)	0.0
Q9VB18	CG5586 protein - Drosophila melanogaster (Fruit fly), 1478 aa.	37..669 86..702	302/650 (46%) 393/650 (60%)	e-150
Q922C2	Unknown (Protein for IMAGE:3592258) - Mus musculus (Mouse), 256 aa (fragment).	1288..1543 1..256	245/256 (95%) 251/256 (97%)	e-136
T15282	hypothetical protein M01D7.3 - Caenorhabditis elegans, 559 aa.	14..495 103..538	151/492 (30%) 231/492 (46%)	1e-51

5

PFam analysis predicts that the NOV74a protein contains the domains shown in the Table 74E.

Table 74E. Domain Analysis of NOV74a

Pfam Domain	NOV74a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Tub	1466..1535	31/86 (36%) 49/86 (57%)	1.5e-27

Example 75.

The NOV75 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 75A.

Table 75A. NOV75 Sequence Analysis			
	SEQ ID NO: 267	1838 bp	
NOV75a, CG140179-01 DNA Sequence	AAATGGCCCAAGAAATAGATCTGAGTGTCTCTCAAGGAGTTAGAACGCGAGGCCATTCTCCAGG TCCTGTACCGAGACCAGGCGGTTCAAACACAGAGGAGGAGAGGACACGGAAACTGAAACAC ACCTGCAGCATCTCCGGTGGAAAGGAGCGAAGAACACGGACTGGGAGCACAAAGAGAAAGTCT GTGCGCGCTGCCAGCAGGTGCTGGGGTTCCTGCTGCACCGGGGCGCCGTGTGCCGGGGTGCA GCCACCGCGTGTGTGCCAGTGCCGAGTGTTCTGAGGGGGACCCATGCCTGGAAGTGCACGG TGTGCTTCGAGGACAGGAATGTCAAAATAAACTGGAGAATGGTCTATGAGGAACGAGCCA AGAAATTTCCAACCTGCAGGCAAAACATGAGACAGTTGGAGGGCAGCTCTTGCAATCTTATCAGA AGCTGAGCAAAATTTCTGTGGTTCCTCTACTCCACCTCCTGTGACGAGAGCCAGTGCAGCC GCAGTAGGCTCCAGGAGTTTGGTCAGTTTAGAGGATTAAATAAGTCCGTGGAAATTTGTTTC TGTCTCTTGCTACCCACGTGAAGAGCTCTCCAAATCCAGAATGATATGACTTCTGAGAAGC ATCTTCTCGCCACGGCCCCAGGCAGTGTGTGGGACAGACAGAGAGACGGAGCCAGTCTGACA CTGCGGTCAACGTCAACACCAGGAAGGTGAGTGCACACAGATATCTGAAACCTCTCAATCAAG AGGATCCCAATGCTCTACTAACCTATTTGAAGCAACAGAATCTCCCATCCAGTCCGGCAC CCAGTACCATATTCTCTGGAGGTTTGTAGACACGGAAGTTTAAATAGCATTGACAGCACCTGTA CAGAGATGGGCAATTTGACAATGCTAATGTCACTGGAGAAATAGAATTTGCCATTCTATTATT GCTTCAAAACCCATTCTTTAGAAATATGCATCAAGGCCTGTAAGAACCTTGCCTATGGAGAAG AAAAGAAGAAAAAGTGAATCCGTATGTGAAGACCTACCTGTTGCCCGACAGATCTCTCCAGG GAAAGCGCAAGACTGGAGTCCAAAGGAACACCGTGGACCCGACCTTTCAGGAGACCTTGAAGT ATCAGGTGGCCCCGTGCCAGCTGGTGACCCGGCAGCTGCAGGTCTCGGTGTGGCATCTGGGCA CGCTGGCCCGGAGAGTGTCTTGGAGAAGTGATCATTCCTCTGGCCACGTGGGACTTTGAAG ACAGCACAAACAGTCTCTCCGCTGGCATCCGCTCCGGGCCAAGGCGGAGAAATACGAAGACA CGCTTCTCAGAGTAATGGAGAGCTCACAGTCCGGGCTAAGCTGGTCTCCCTTCACGGCCCCA GAAACTCCAAGAGGCTCAAGAAGGTGAGCCATCACTTTCATGGTCACTTTGTTGGTAGTGC TAGGAGCCAAGAATTACCTGTGCGGCCAGATGGCACCTTGAACCTATTGTTAAGGGGTGTC TCACTCTGCCAGACCAACAAAACCTGAGACTGAAGTCGCCAGTCTGAGGAAGCAGGCTTGCC CCCAGTGGAAACACTCATTGTCTTCACTGGCGTAACCCAGCTCAGCTGAGGCAGTCAAGCT TGGAGTTAACTGTCTGGGATCAGGCCCTCTTGGAAATGAACGACCGCTTGTGGAGGAACCA GACTTGGTTCAGAGGGAGACACAGCTGTTGGCGGGATGCATGCTCACTATCGAAGCTCCAGT GGCAGAAAGTCTTCCAGCCCCAATCTATGGACAGACATGACTCTTGTCTGCACTGACATG AAGGCCTCAAG		
	ORF Start: ATG at 3		ORF Stop: TGA at 1821
	SEQ ID NO: 268	606 aa	MW at 68204.5kD
NOV75a, CG140179-01 Protein Sequence	MAQEIDLSALKELEREAILQVLYRDQAVQNTEERTRKLKTHLQHLRWKGAKNTDWEHKEKCC ARCQQVLGFLLRGAVCRGCSHRVCAQCRVFLRGTHAWKCTVCFEDRNVKIKTGEWFYEERAK KFPTAGKHETVGGQLQSYQKLSKISVVPPTPPPVSSESQCSRSRLQEFQGRFGFNKSVENLFL SLATHVKELSKSQNDMTSEKHLATGPRQCVGQTERRSQSDTAVNVTRKVSAPDILKPLNQE DPKCSTNPILKQQLPSSPAPSTIFSGGFRHGSLSISIDSTCTEMGNFDNANVTGEIEFAIHYC FKTHSLEICIKACKNLAYGEEKKKCNPYVKTYLLPDRSSQGRKTGVQRNTVDPTFQETLKY QVAPAQLVTRQLQVSVVHLGLTARRVFLGEVILPLATWDFEDSTTQSFWRHPLRAKAEKYEDS VPQSNGETVRAKLVLPSRPRKLQEAQEGQPSLHGQLCLVVLGAKNLPVRPDGTLNSFVKGCL TLPDQQLRLKSPVLRKQACQWKHSFVFSGVTPAQLRQSSLELTVDQALFGMNDRLLGSTR LGSEGDVAVGGDACSLSKLQWQVLSSPNLWDTMTLVLH		

5

Further analysis of the NOV75a protein yielded the following properties shown in Table 75B.

Table 75B. Protein Sequence Properties NOV75a	
PSort analysis:	0.6000 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV75a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 75C.

5

Table 75C. Geneseq Results for NOV75a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV75a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAG67216	Amino acid sequence of human Parkin-Associated Protein 1 (PAP1) - Homo sapiens, 610 aa. [WO200160857-A2, 23-AUG-2001]	1..606 1..610	602/610 (98%) 604/610 (98%)	0.0
AAG67214	Amino acid sequence of human Parkin-Associated Protein 1 (PAP1) - Homo sapiens, 610 aa. [WO200160857-A2, 23-AUG-2001]	1..606 1..610	602/610 (98%) 604/610 (98%)	0.0
AAU19743	Human novel extracellular matrix protein, Seq ID No 393 - Homo sapiens, 363 aa. [WO200155368-A1, 02-AUG-2001]	144..502 1..363	352/363 (96%) 353/363 (96%)	0.0
AAU87138	Novel central nervous system protein #48 - Homo sapiens, 363 aa. [WO200155318-A2, 02-AUG-2001]	144..502 1..363	352/363 (96%) 353/363 (96%)	0.0
AAG67212	Amino acid sequence of human Parkin-Associated Protein 1 (PAP1) - Homo sapiens, 344 aa. [WO200160857-A2, 23-AUG-2001]	265..606 1..344	340/344 (98%) 341/344 (98%)	0.0

In a BLAST search of public sequence databases, the NOV75a protein was found to have homology to the proteins shown in the BLASTP data in Table 75D.

Table 75D. Public BLASTP Results for NOV75a				
Protein Accession Number	Protein/Organism/Length	NOV75a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99N54	Synaptotagmin-like protein 3-a - Mus musculus (Mouse), 607 aa.	1..606 1..607	485/607 (79%) 543/607 (88%)	0.0
CAC69571	Sequence 1 from Patent WO0160857 - Homo sapiens (Human), 344 aa (fragment).	265..606 1..344	340/344 (98%) 341/344 (98%)	0.0
Q99N49	Synaptotagmin-like protein 3-a + 3S-I (Synaptotagmin-like 3) - Mus musculus (Mouse), 412 aa.	195..606 1..412	332/412 (80%) 370/412 (89%)	0.0
Q99N48	Synaptotagmin-like protein 3-a delta 3S-II - Mus musculus (Mouse), 393 aa.	216..606 3..393	316/391 (80%) 351/391 (88%)	0.0
Q99N79	Synaptotagmin-like protein 3-b - Mus musculus (Mouse), 393 aa.	216..606 3..393	315/391 (80%) 351/391 (89%)	0.0

PFam analysis predicts that the NOV75a protein contains the domains shown in the Table 75E.

Table 75E. Domain Analysis of NOV75a			
Pfam Domain	NOV75a Match Region	Identities/ Similarities for the Matched Region	Expect Value
RPH3A_effector	1..252	62/326 (19%) 117/326 (36%)	0.27
C2	321..410	26/97 (27%) 64/97 (66%)	1.3e-11
C2	478..567	27/98 (28%) 62/98 (63%)	0.0029

5

Example 76.

The NOV76 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 76A.

Table 76A. NOV76 Sequence Analysis			
	SEQ ID NO: 269	1545 bp	
NOV76a,	TACTGAGGCTTTCGGGACGGCGCGGGAAGATGGCGGCCTCCAGGAATGGGTTTGAAGCCGTG GAGGCAGAGGGCAGCGCAGGGTCCCGGGAAGCTCGGGAATGGAGGTGGTCTTCTTTGGAT		

CGI40392-01 DNA Sequence	CCTGCCGTCCCCGCCCGCTGTGCCCTCACGGTCCCACTCTTCTGTTTGTAAAGGTGACCCAA GGGAAAGAAGAACTCGGAGGTTTTATGCCTGTTAGAGATAGAAAAGACTGTAAT TTTTTTCAGTGGGAAGATGAAAAGTTGTCAGGAGCTAGACTTGCTGCCCGAGAAGCTCATAAC CGAAGATGTCAGCCTCCCCGTGCCGAACGCAGTGTGTGGAAAGGTACTTGAAGTTTATTGAG TTGCCCTTGACTCAGAGAAAGTTTGTCAAACATGTCAGCAGTTGTTGTTACAGATGACTGG GGGCAACATAGTGAGCATCAGGTTCTGGGTAATGTGTCTCATTACCCAGTTAAGAAGGCCAGT CAACTCCTTTATCCACTGGAACAAGAAGACAAATGCCAGTATCTGTTTGCTGATCGGAGC TGTCAGTTCTTGGTAGACTTACTTTCTGCCCTCGGATTGAGAAGAGTACTGTGTGTTGGAACA CCAAGGTTGCATGAGCTGATCAAGTTGACAGCATCAGGTGACAAGAAGTCTAACATTAAAAAGC CTTTTATTGGATATTGATTTTAGGTATTCACAGTTTATATGGAAGATAGCTTTTGCCATTAT AATATGTTTAACCATCATTTCTTTGATGGAAGAAGCTGCCCTTGAAGTATGCAGAGCATTTTAT CAGGAAGATAAAGGCGAAGGAATCATTATGGTTACGGATTCTCCGTTTGGTGGCTTGGTTGAA CCTCTGGCTATTACATTCAAGAAGTTAATTGCTATGTGGAAGAAGGTCAAAGCCAAGATGAC AGTCACAAGAAGTACCCATTTTCTGGATTTTCCCTATTTTGAATCCCGAATTGTCAG TTTTTTCAGCTTCCAGATGCTGGATTACAGGTAGATTATGATAATCATGCACTTTATATAA CACGGAAGACAGGTGAAAAACAGTCTCCCGTGCCTATTTTACCAACATTCCGCCCAACAAA ATAATCCTTCTCTACTGAAGAAGGGTACAGGTTTGTCTCCGTGTCAACGGTATGTTTCTCTA GAGAATCAACACTGTGAGCTCTGTAATCTTGCACATCCAAGGATGGCAGGAAATGGAACCAT TGCTTCTCTGTGTAAGAAGTGTGTAAGCCTGCCTGGATCCACTGTAGCATCTGCAATCACTGT GCTGTTCCAGATCATTTCTGTGAGGGCCCCAACATGGCTGCTTTATTTGTGGTGAAGTGGAT CATAAACGCAGTACTTGTCTTAACATTGCTACATCTAAGAGAGCTAACAAAGTCAGTCAAAAA AAAAAAAAAAAAAAAAAAAAAGCCACAGAGAAAAGAAAGGGCACCACCGGTAGGTTTGG AATCAGAAAATTTGGCAAGAAGAGATTGAAA		
	ORF Start: ATG at 31		ORF Stop: TAG at 1504
	SEQ ID NO: 270	491 aa	MW at 56449.4kD
NOV76a, CGI40392-01 Protein Sequence	MAASRNGFEAVEAEGSAGCRGSSGMEVVLPLDPAVPAPLCPHGPTLLFVKVTQKKEETRRFYA CSACRDRKDCNFFQWEDEKLSGARLAAREAHNRRCQPPLSRTQCVERYLKFIELPLTQRKFCQ TCQQLLLPDDWQHQSEHQVLGNVSTQLRRPSQLLYPLENKKTNAYLFADRSCQFLVDLLSA LGFRRLVCVGTPLRHELIKLTASGDKSNIKSLLLDIDFRYSQFYMEDSFCHYNMFNHHFFDQ KTALEVCRFLQEDKGEIIMVTDSPFGGLVEPLAITFKKLIAMWKEGQSQDDSHKELPIFWI FPYFFESRICQFFPSQMLDYQVDYDNHALYKHGKTGRKQSPVRIFTNIPPNKIILPTEEGYR FCSPCQRYVSLNQHCCLNSCTSKDGRKWNHCFCKKCVKPAWIHCSCINHCAPVDHSCCEGP KHGCFICGELDHKRSTCPNIATSKRANKSVKKKKKKKKKSHREKKGHHR		

Further analysis of the NOV76a protein yielded the following properties shown in Table 76B.

Table 76B. Protein Sequence Properties NOV76a	
PSort analysis:	0.9800 probability located in nucleus; 0.3725 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV76a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 76C.

Table 76C. Geneseq Results for NOV76a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV76a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU28264	Novel human secretory protein, Seq ID No 621 - Homo sapiens, 328 aa. [WO200166689-A2, 13-SEP-2001]	1..303 11..314	264/304 (86%) 271/304 (88%)	e-153
AAU28076	Novel human secretory protein, Seq ID No 245 - Homo sapiens, 237 aa. [WO200166689-A2, 13-SEP-2001]	1..233 1..233	233/233 (100%) 233/233 (100%)	e-136
ABB67501	Drosophila melanogaster polypeptide SEQ ID NO 29295 - Drosophila melanogaster, 349 aa. [WO200171042-A2, 27-SEP-2001]	356..491 196..331	53/139 (38%) 77/139 (55%)	2e-25
ABB68484	Drosophila melanogaster polypeptide SEQ ID NO 32244 - Drosophila melanogaster, 968 aa. [WO200171042-A2, 27-SEP-2001]	389..431 194..234	20/43 (46%) 23/43 (52%)	0.003
ABB61064	Drosophila melanogaster polypeptide SEQ ID NO 9984 - Drosophila melanogaster, 285 aa. [WO200171042-A2, 27-SEP-2001]	408..438 105..136	15/32 (46%) 20/32 (61%)	0.050

In a BLAST search of public sequence databases, the NOV76a protein was found to have homology to the proteins shown in the BLASTP data in Table 76D.

Table 76D. Public BLASTP Results for NOV76a

Protein Accession Number	Protein/Organism/Length	NOV76a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D2I1	4930449I23Rik protein - Mus musculus (Mouse), 489 aa.	1..485 1..483	388/485 (80%) 431/485 (88%)	0.0
Q9H5U6	CDNA: FLJ23024 fis, clone LNG01684 - Homo sapiens (Human), 279 aa.	235..485 1..251	239/251 (95%) 245/251 (97%)	e-151
Q96AN7	Hypothetical 23.6 kDa protein - Homo sapiens (Human), 205 aa.	25..229 1..205	205/205 (100%) 205/205 (100%)	e-119
Q9NZY3	HSPC052 - Homo sapiens (Human), 127 aa.	25..144 1..119	116/120 (96%) 117/120 (96%)	2e-65
AAF58228	CG12863-PA - Drosophila melanogaster (Fruit fly), 425 aa.	32..491 4..407	150/472 (31%) 225/472 (46%)	2e-62

PFam analysis predicts that the NOV76a protein contains the domains shown in the Table 76E.

Table 76E. Domain Analysis of NOV76a			
Pfam Domain	NOV76a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-CCHC	443..460	6/18 (33%) 11/18 (61%)	0.045

5

Example 77.

The NOV77 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 77A.

Table 77A. NOV77 Sequence Analysis			
	SEQ ID NO: 271	1945 bp	
NOV77a, CG140727-01 DNA Sequence	CGAAGGAGGTGGTGGCTGCGTTGGGCTCCGGGAAGCCGTTCCGGCTGGGGCTGTCGGCCGCGG GGCGGAGGCACTCGCGCGGGGGTAATTCCGGGTTCTGGGTCCGCGCAGCTTTCCCC GGATTGACTTGGCCTCTACTTCTTGTAAAGGAAATTCATCTTGTTTATCAGGTGTGTGTG GTTTCAGCGCAGCATGGCTGTGGTCATCCGTTTGCAAGGCTCCCAATTGTGGCGGGGACCAT GGACATTCGCCACTTCTTCTCTGGATTGACCATTCTGATGGGGCGTGCATATTGTAGGGGG TGAAGTGGATGGCCCACTGCGTGACCTTGGTTCAGCTGTCCATTTCTGTGACCATCTCATTG ACAAGGACATCGGCTCCAAGTCTGACCACTCTGCGTCCTTTTACAGGATGTGGGAGGGGGCA GCTGGGCTGAGCTTGGCCGGAAGTGAACGGGTGCGGAAGTCTCAAGCCCTGAGTTCTCCAAGA CTCTACAGCTTGAGTACCGCTTTGAGACAGTCCAGAAGCTACGCTTTGGAATCTATGACATAG ACAACAAGACGCCAGAGCTGAGGGATGATGACTTCTAGGGGGTGTGAGTGTTCCTTAGGAC AGATTGTGTCCAGCCAGGTACTGACTCTCCCTTGATGCTGAAGCCTGGAACCTGCTGGGC GGGGACCATCACGGTCTCAGTCAAGGAATTAAGGACAATCGTGTAGTAACCATGGAGGTAG AGGCCAGAAACCTAGATAAGAAGGACTTCTGGGAAAATCAGATCCATTTCTGGAGTTCTTCC GCCAGGTGATGGAAATGGCACCTGGTGTACAGATCTGAGGTATCAAGAACAACCTGAACC CTACATGGAAGCGTTTCTCAGTCCCGGTTGAGCATTCTGTGGTGGGAACCCAGCACACCCA TCCAGGTGCAATGCTCCGATTATGACAGTGACGGGTACATGATCTCATCGGTACCTTCCACA CCAGCTTGGCCCACTGTCAGGCAGTCCCGGCTGAGTTTGAATGCATCCACCTGAGAAGCAGC AGAAAAAGAAAAGCTACAAGAACTCTGGAAGTATCCGTGTCAAGATTGTGGGTAGAAACAG AGTACTCCTTTCTGGACTATGTGATGGGAGGCTGTGAGTCAACTTCACTGTGGCGTGGACT TCACTGGCTCCAATGGAGACCCCTCCTCACCTGACTCCCTACACTACCTGAGTCCAACAGGGG TCAATGAGTACCTGATGGCACTGTGGAGTGTGGGAGCGTGGTTCAGGACTATGACTCAGACA AGCTGTTCCCTGCATTTGGATTGGGGCCAGGTTCCCCCTGACTGGCAGGTCTCGCATGAAT TTGCCTTGAATTTCAACCCAGTAACCCCTACTGTGCAGGCATCCAGGGCATTGTGGATGCCT ACCGCCAAGCCCTGCCCAAGTTGCGCTCTATGGCCCTACCAACTTGCACCCATCATCAACC ATGTGGCCAGGTTTGCAGCCAGGCTGCACATCAGGGGACTGCCTCGCAATACTTCATGCTGT TGCTGCTGACTGATGGTGTGTGACGGATGTGGAAGCCACACGTGAGGCTGTGGTGCCTGCCT CGAACCTGCCATGTGAGTATCATTGTGGGTGTGGGTGGTGTGACTTTGAGGCCATGGAGC AGCTGGACGCTGATGGTGGACCCCTGCATACACGTTCTGGGAGGCTGCTGCCCGGACATTG TGCAGTTTGTACCCTACCGCGGTTCCAGAATGTGAGTAGGTGCAAAATTTGATCTGGGAGTTT ACCTTTACCTTTCCCTGATGTGAGTAGGGTTGAGGCCAAGGGATGCTCACCTTTTCTCT AAGCCTGGGATATGGTTGGGTTCTTGTAGCATGAAATACAGACTTTGGGCAGCTGT		
	ORF Start: ATG at 324		ORF Stop: TGA at 1845
	SEQ ID NO: 272	507 aa	MW at 56052.8kD
NOV77a,	MAHCVTLVQLSISCDHLIDKDIGSKSDPLCVLLQDVGGGSWAELGRTERVRNCSSPEFSKTLQ		

CG140727-01 Protein Sequence	LEYRFETVQKLRFGIYDIDNKTPELRDDDFLGGAECSLGQIVSSQVLTLPMLKPGKPAGRGT ITVSAQELKDNRRVVTMEVEARNLDKKDFLGKSDPFLEFFRQGDGKWHLYRSEVIKNNLNPTW KRFSVPVQHFCGGNPSTPIQVQCSYDSDGSHDLIGTFHTSLAQLQAVPAEFECIHPEKQKK KSYKNSGTIRVKICRVETEYSFLDYVMGGCQINFTVGVDFTGSNGDPSSPDSLHYLSPTGVNE YLMALWSVGSVVQDYSDKLFPAFGGAQVPPDWQVSHEFALNFNPSNPYCAGIQGIVDAYRQ ALPQVRLYGPTNFAPIIINHVARFAAQAAHQGTASQYFMLLLTDGAVTDVEATREAVVRASNL PMSVIVGVGGADFEAMEQLDADGGPLHTRSGQAAARDIVQFVPRRFQNVSRCKFDLGVPFTFP
------------------------------	--

Further analysis of the NOV77a protein yielded the following properties shown in Table 77B.

Table 77B. Protein Sequence Properties NOV77a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3359 probability located in microbody (peroxisome); 0.1756 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	Cleavage site between residues 14 and 15

5

A search of the NOV77a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 77C.

Table 77C. Geneseq Results for NOV77a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV77a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAU19736	Human novel extracellular matrix protein, Seq ID No 386 - Homo sapiens, 540 aa. [WO200155368-A1, 02-AUG-2001]	2..494 6..503	331/498 (66%) 404/498 (80%)	0.0
AAM39997	Human polypeptide SEQ ID NO 3142 - Homo sapiens, 548 aa. [WO200153312-A1, 26-JUL-2001]	3..494 21..516	294/497 (59%) 378/497 (75%)	0.0
AAB24231	Human vesicle associated protein 10 SEQ ID NO:10 - Homo sapiens, 532 aa. [WO200060082-A2, 12-OCT-2000]	3..494 5..500	294/497 (59%) 378/497 (75%)	0.0
AAG62450	Human membrane-binding protein 37 - Homo sapiens, 336 aa. [WO200138363-A1, 31-MAY-2001]	1..310 30..333	302/310 (97%) 303/310 (97%)	e-178

AAU19854	Human novel extracellular matrix protein, Seq ID No 504 - Homo sapiens, 399 aa. [WO200155368-A1, 02-AUG-2001]	57..440 6..394	259/389 (66%) 318/389 (81%)	e-160
----------	---	-------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV77a protein was found to have homology to the proteins shown in the BLASTP data in Table 77D.

Table 77D. Public BLASTP Results for NOV77a				
Protein Accession Number	Protein/Organism/Length	NOV77a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99829	Copine I - Homo sapiens (Human), 537 aa.	1..494 1..494	492/494 (99%) 492/494 (99%)	0.0
Q925K4	Copine I protein - Mus musculus (Mouse), 454 aa (fragment).	1..452 1..451	417/452 (92%) 436/452 (96%)	0.0
Q925K5	Copine I protein - Mus musculus (Mouse), 448 aa (fragment).	1..449 1..448	414/449 (92%) 433/449 (96%)	0.0
O75131	Copine III - Homo sapiens (Human), 537 aa.	2..494 3..500	332/498 (66%) 405/498 (80%)	0.0
Q96FN4	Unknown (protein for MGC:16924) - Homo sapiens (Human), 446 aa.	88..494 3..414	255/412 (61%) 325/412 (77%)	e-158

5

PFam analysis predicts that the NOV77a protein contains the domains shown in the Table 77E.

Table 77E. Domain Analysis of NOV77a			
Pfam Domain	NOV77a Match Region	Identities/ Similarities for the Matched Region	Expect Value
C2	7..98	22/100 (22%) 60/100 (60%)	0.036
C2	140..228	30/101 (30%) 63/101 (62%)	5.4e-06

10

Example 78.

The NOV78 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 78A.

Table 78A. NOV78 Sequence Analysis

	SEQ ID NO: 273	435 bp	
NOV78a, CG141070-01 DNA Sequence	GGGCAAGTGCACAGTGGTCTGGCGGCCCATGTCATTCTGCAGCTTCTCGGGGGCAAGGTT TTCCAGAATCACTTTGAGCCAGGCGTCTATATGTGTGCCAAGTGTGGCTATGAGCTGTTTCCC AGCCGCTCAAAGTACACATACTCATTCCCCTGGCCGGTGTTACCAAGACCATCCGTTCTGAC AGCGTGGCCAAGCGCCAGAGCACAATCATCCTGAAAGTCTTGAAGGTGTCTCGTGCAAGTGT GGCAACACGTTGAGCCACAAGTTCCTGAACGATGGCCCCAAGCTGCGGCAGTCCCGATTATA TTCAGCAGCTCGCTGAAGTTTGTCCCTAAAGGCAAGAACTTCTGCTTCCAGGCGCACTAG GCGGGCAGCCACACCGACCCAGATGGCCACGGCACTAAGGCCACACACTGGCCAT		
	ORF Start: ATG at 31		ORF Stop: TAG at 376
	SEQ ID NO: 274	115 aa	MW at 13039.8kD
NOV78a, CG141070-01 Protein Sequence	MSFCSFFGGKVFQNHFEFPGVYMCACGYELFPSRSKYTYSFPWPVFTKTIRSDSVAKRPEHNNH PESLEGVSCCKGNTLSHKFLNDGPKLRQSRFIFSSSLKFVPKGKETSASQAH		

Further analysis of the NOV78a protein yielded the following properties shown in Table 78B.

5

Table 78B. Protein Sequence Properties NOV78a	
PSort analysis:	0.6400 probability located in microbody (peroxisome); 0.4500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV78a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 78C.

10

Table 78C. Geneseq Results for NOV78a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV78a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB37402	Human secreted protein BLAST search protein SEQ ID NO: 112 - Homo sapiens, 99 aa. [WO200058335-A1, 05-OCT-2000]	7..104 1..99	78/100 (78%) 85/100 (85%)	6e-39
AA Y60509	Human normal bladder tissue EST encoded protein 181 - Homo sapiens, 138 aa. [DE19818620-A1, 28-OCT-1999]	1..94 45..138	74/95 (77%) 82/95 (85%)	8e-39
AAW46757				4e-32

	[WO9748797-A1, 24-DEC-1997]	1..83	71/84 (84%)	
ABG34053	Human Pro peptide #24 - Homo sapiens. 192 aa. [WO200224888-A2, 28-MAR-2002]	5..111 69..175	36/109 (33%) 56/109 (51%)	8e-09
AAM41523	Human polypeptide SEQ ID NO 6454 - Homo sapiens, 201 aa. [WO200153312-A1, 26-JUL-2001]	5..111 78..184	36/109 (33%) 56/109 (51%)	8e-09

In a BLAST search of public sequence databases, the NOV78a protein was found to have homology to the proteins shown in the BLASTP data in Table 78D.

Table 78D. Public BLASTP Results for NOV78a				
Protein Accession Number	Protein/Organism/Length	NOV78a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96RX6	Annexin A2 like-? : selenoprotein X - Homo sapiens (Human), 116 aa.	1..115 1..116	94/117 (80%) 102/117 (86%)	1e-48
Q9NZV6	Selenoprotein X 1 (Protein HSPC270) - Homo sapiens (Human), 116 aa.	1..115 1..116	94/117 (80%) 102/117 (86%)	1e-48
Q9JLC3	Selenoprotein X 1 (Selenoprotein R) - Mus musculus (Mouse), 116 aa.	1..115 1..116	88/117 (75%) 96/117 (81%)	1e-44
Q9BTV2	Similar to selenoprotein X, 1 - Homo sapiens (Human), 94 aa.	1..94 1..94	74/95 (77%) 82/95 (85%)	2e-38
AAM31330	Transcriptional regulator - Methanosarcina mazei (Methanosarcina frisia), 140 aa.	7..108 34..138	40/107 (37%) 57/107 (52%)	3e-09

5

PFam analysis predicts that the NOV78a protein contains the domains shown in the Table 78E.

Table 78E. Domain Analysis of NOV78a			
Pfam Domain	NOV78a Match Region	Identities/ Similarities for the Matched Region	Expect Value
DUF25	12..63	18/52 (35%) 34/52 (65%)	0.00072

Example 79.

The NOV79 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 79A.

Table 79A. NOV79 Sequence Analysis			
	SEQ ID NO: 275	1044 bp	
NOV79a, CG141395-01 DNA Sequence	CACGTCCTCCCTGTGCGGCCAGCGTCAGAGCCATGGCGATGGAGGAGAGGAAGCCCGAGACCGAG GCAACGAGAGCACAGCCGACCCCTTCGTCATCCACCACTCAGAGCAAGCCTACGCCCGTGAAG CCAAACTATGCTCTCAAGTTCACCCCTTGCTGGCCACACCAAGCAGTGTCCTCCGTGAAATTC AGCCCGAATGGAGAGTGGCTGGCAAGTTCATCTGCTGATAAACTCATTAAATTTGGGGGTCA TATGATGGGAAATTTGAGAAAACCATGTCTGGTCACAGCCTGTGGTCGTAGATTCTAACCTT TTGTTTCCGCCTCAGATGACAAAACCTTGAAGATACGGGACGTGAGCTCGGGAAAGTGTCTG AAAACCTGAAGGGACACAGTAATTATGTCTTTGCTGTAACTTCAATCCCCAGTCCAGCCTT ACTGTCTCAGGATCCTTTGATGAAAGTGTGAGGATATGGGTTGTGAAAACAGGGAAGTGCCAC AAGACTCTGCTAGCTCACTCCGATCCAGTCTCGGCCATTCAATTTAATCGTGATGGATCTTG ATAGTTTCAAGTAGCTATGATGGTCTCTGTACATCTGGGACACCGCCTCAGGCCAGTGCCCTG AAAACGCTCACTGATGATGACAACCCCTGGTGTCTTTCGTGAAGCTCTCCCGAAGGGTGGA TACATCGTGGCTGCCACGCTGGGCAACACACTCAAGCTCTGGGACTACAGCAAGGGGAAGTGC CTGAAGACATACACTGGCCACAAGAACGAGAAATACTGCATATTTGCTAATTTCTCTGTACT GGCGGGAAGTGGATTGTGTCTGGCTCGGAGGATAACCTTCTTACATCTGGAAACTTCAGACG AAAGAGATTGTACAGAAATTAGAAGGCCACACAGATGTTGTGACCTCAACAGCTTGTCACCCA ACAGAAACATCATCACCTCTGCCGCGCTAGAAAATGACAAAACAATTAAACTGTGGAAGAGT GACTGTAAAGTCCCTTTGCTCCCATGCGATAGAC		
	ORF Start: ATG at 31		ORF Stop: TAA at 1015
	SEQ ID NO: 276	328 aa	MW at 36088.6kD
NOV79a, CG141395-01 Protein Sequence	MAMEERKPEATEATRAQPTPSSSTTQSKPTPVKPNYALKFTLAGHTKAVSSVKFSPNGEWLASS SADKLIKIWGSYDGKFEKTMGSHLSWSSDSNLFVSASDDKTLKIRDVSSGKCLKTLKGHSNYV FCCNFPQSSLTVSGSFDSESVRIWVKTGKCHKTLAHSDFVSAIHFNRDGFLIVSSSYDGLC HIWDTASGQCLKTLTDDDNPLVSFVKLSPKGGYIVAATLGNTLKLWDYSKGLKTYTGHKNE KYCIFANFSVTGGKWIVSGSEDNLLYIWKLTKEIVQKLEGHTDVVTSTACHPTENIITSAL ENDKTIKLWSDC		

5

Further analysis of the NOV79a protein yielded the following properties shown in Table 79B.

Table 79B. Protein Sequence Properties NOV79a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.4206 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV79a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 79C.

Table 79C. Geneseq Results for NOV79a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV79a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB97345	Novel human protein SEQ ID NO: 613 - Homo sapiens, 334 aa. [WO200222660-A2, 21-MAR-2002]	1..328 1..334	299/334 (89%) 307/334 (91%)	e-176
AAB68529	Human GTP-binding associated protein #29 - Homo sapiens, 334 aa. [WO200105970-A2, 25-JAN-2001]	1..328 1..334	299/334 (89%) 307/334 (91%)	e-176
AAB63186	Human secreted protein sequence encoded by gene 3 SEQ ID NO:112 - Homo sapiens, 317 aa. [WO200061629-A1, 19-OCT-2000]	17..327 1..317	285/317 (89%) 292/317 (91%)	e-168
ABB68576	Drosophila melanogaster polypeptide SEQ ID NO 32520 - Drosophila melanogaster, 361 aa. [WO200171042-A2, 27-SEP-2001]	15..327 42..360	260/319 (81%) 279/319 (86%)	e-154
AAB93659	Human protein sequence SEQ ID NO:13175 - Homo sapiens, 330 aa. [EP1074617-A2, 07-FEB-2001]	9..327 5..329	253/325 (77%) 275/325 (83%)	e-149

In a BLAST search of public sequence databases, the NOV79a protein was found to have homology to the proteins shown in the BLASTP data in Table 79D.

Table 79D. Public BLASTP Results for NOV79a				
Protein Accession Number	Protein/Organism/Length	NOV79a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9UGP9	WD-repeat protein 5 (WD repeat protein BIG-3) - Homo sapiens (Human), and, 334 aa.	1..328 1..334	299/334 (89%) 307/334 (91%)	e-176
Q8T776	Hypothetical 38.6 kDa protein - Branchiostoma floridae (Florida lancelet) (Amphioxus), 353 aa.	1..327 1..352	277/352 (78%) 298/352 (83%)	e-160
Q9V3J8	Will die slowly protein - Drosophila melanogaster (Fruit fly), 361 aa.	15..327 42..360	260/319 (81%) 279/319 (86%)	e-153
Q9NUL4	CDNA FLJ11287 fis, clone PLACE1009596, weakly similar to vegetatible incompatibility protein HET-E-1 - Homo sapiens (Human), 330 aa.	9..327 5..329	253/325 (77%) 275/325 (83%)	e-148

Q9D7H2	2310009C03Rik protein - Mus musculus (Mouse), 328 aa.	12..328 8..328	235/323 (72%) 271/323 (83%)	e-139
--------	---	-------------------	--------------------------------	-------

PFam analysis predicts that the NOV79a protein contains the domains shown in the Table 79E.

Table 79E. Domain Analysis of NOV79a

Pfam Domain	NOV79a Match Region	Identities/ Similarities for the Matched Region	Expect Value
WD40	37..73	17/37 (46%) 30/37 (81%)	4.3e-06
WD40	115..151	17/37 (46%) 31/37 (84%)	8.2e-06
WD40	157..193	15/37 (41%) 29/37 (78%)	9.4e-06
WD40	242..281	10/40 (25%) 32/40 (80%)	0.04
WD40	287..325	12/39 (31%) 29/39 (74%)	0.18

5

Example 80.

The NOV80 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 80A.

Table 80A. NOV80 Sequence Analysis

	SEQ ID NO: 277	1602 bp
NOV80a, CG191018-01 DNA Sequence	GGAGCCATGCGGCGATCGAGGAGCTCTGCGGCCGCCAAGCTGCGCGGGCAGAAAGCGGTCCGGG GCCTCCGCGGCCCGCGGGCCTCCGCGGCCGTGCTTGGCACCCAGCGCCACCCGCACACGG CGCTCCGCTAGCCAGGCCGGGAGCAAGAGCCAGGCGGTGGAGAAGCCGCGCTCGGAGAAGCCG CGGCTGAGGCGCTCGTCGCGCGGGGCCAGGAGGAGGCCCGGGGAGCCGCGCCGCTGAG CTGGCGTTGCTCCCGCCACCGCCGCGCCGCGCCGACTCCCGCGACCCGACGTCCTCGGCG TCCAACCTGGACCTGGGCGAGCAGCGGGAGCGCTGGGAGACGTTCCAGAAGCGGCAGAAGCTT ACCTCCGAGGGTGCCGCCAAGCTCCTGCTAGACACCTTTGAATACCAGGGCCTGGTGAAGCAC ACAGGAGGCTGCCACTGTGGAGCAGTTCGTTTGAAGTTTGGGCCTCAGCAGACTTGCATATA TTTGACTGCAATTGCAGCATTTGCAAGAAGAAGCAGAATAGACACTTCATTGTTCCAGCTTCT CGCTTCAAGCTCCTGAAGGAGCTGAGCACATAACGACTTACAGTTCAATACTCACAAAGCC CAGCATACCTTCTGTAAGAGATGTGGCGTTCAGAGCTTCTATACTCCACGATCAAACCCGGA GGCTTCGGAATTGCCCCCACTGCCTGGATGAGGGCACTGTGCGGAGTATGGTCACTGAGGAA TTCAATGCGAGCGATTGGGAGAAGGCCATGAAAGAGCACAAAGACCATCAAGAACATGTCTAAA GAGTGAGCTTCTGCCTCTCTGCCTGAAAAGGAGGAATGATTGGGGCCAGCAACTTTGCTCT CCCTGCCGTGCCTCGGTGGTGTCTCTGAATGTGGCTGACCTGGGCTGCTGGTTCCGTTGACTA GGGTCACTCTGATCTCTGCAGTTTGCTCCAGCTACCAGTTTCTTTAGGCAGCTCTTTGCTCTC CCTCTGCCAGATTTGATGTAGTCTAATGACATCCTTCTTCCCACTTTGTGTGATCC AGCAGAGCATGTGAGACTTTTGATATGCACCTTCATGTATTATCTTGTTCAGTTCTCTGAGG TTGGGATCATTATTATTTCCATTGTCAGATGAGAGAATTGAGGCAGAGAAAGGTTTCAGCAC CTTGCTTTGGTTACACAGCTGGTCATTCTGGCTTCAATCGCAGGACTACCAGCCTGTGCTCT	

	TCACCACTTAGCTTCCCTGACTCAGGCCACTTCCCTGGAGCGTAGCTGGATTCTGAGAGTAG TTTCCAAGCCAGAGCTTTCAGAGAGCTTTTGTTCGTAGGACAATTTTAAGACATCAGGTTCTT GAATGTTTTGTGTTTTTTAAGTCTCAGATTATCTTCTACTTCTCTCTCCAAAAGAC TGAGAGCTGACATATTTGATTGTAAGCTCTTTGAGGCAGAGTTCTTGTAATCGTCTCTGTATA AAACAGTGCCACCCAGTGACCTGTACTTGGATGCTTCAATCAGAGCTGTCTGTTAAATAG AGCAAGTTTTCTTAGACCCACATTCT		
	ORF Start: ATG at 7		ORF Stop: TGA at 823
	SEQ ID NO: 278	272 aa	MW at 29730.3kD
NOV80a, CG191018-01 Protein Sequence	MRRSRSSAAAKLRGQKRSASAAPASAAAALAPSATRRRSASQAGSKSQAVEKPPSEKPR RRSSPRAQEEGPGEPPELALLPPPPPPPTPATPTSSASNLDLGEQERWETFQKRQKLT EGAAKLLDTEYQGLVKHTGGCHGAVRFEVWASDLHIFDCNCSICKKKQNRHFIVPASRF KLLKGAEHITTYTFNTHKAHTFCKRCGVQSFTYTPRNPGGFGIAPHCLDEGTVRSMVTEEFN GSDWEKAMKEHKTIKNMSKE		

Further analysis of the NOV80a protein yielded the following properties shown in Table 80B.

Table 80B. Protein Sequence Properties NOV80a	
PSort analysis:	0.7941 probability located in mitochondrial matrix space; 0.6305 probability located in mitochondrial intermembrane space; 0.4722 probability located in mitochondrial inner membrane; 0.4722 probability located in mitochondrial outer membrane
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV80a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 80C.

Table 80C. Geneseq Results for NOV80a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV80a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB43357	Human ORFX ORF3121 polypeptide sequence SEQ ID NO:6242 - Homo sapiens, 245 aa. [WO200058473-A2, 05-OCT-2000]	32..272 5..245	234/241 (97%) 235/241 (97%)	e-140
AAM38640	Human colorectal cancer antigen SEQ ID NO: 155 - Homo sapiens, 194 aa. [WO200155350-A1, 02-AUG-2001]	79..272 1..194	187/194 (96%) 187/194 (96%)	e-111
AAM93169	Human digestive system antigen SEQ ID NO: 2518 - Homo sapiens, 194 aa. [WO200155314-A2, 02-AUG-2001]	79..272 1..194	187/194 (96%) 187/194 (96%)	e-111
ABB77648	Ribosomal protein s13.09 amino acid	169..272 15..118	103/104 (99%) 104/104 (99%)	5e-59

	[CN1326940-A, 19-DEC-2001]			
AAM78645	Human protein SEQ ID NO 1307 - Homo sapiens, 118 aa. [WO200157190-A2, 09-AUG-2001]	169..272 15..118	103/104 (99%) 104/104 (99%)	5e-59

In a BLAST search of public sequence databases, the NOV80a protein was found to have homology to the proteins shown in the BLASTP data in Table 80D.

Table 80D. Public BLASTP Results for NOV80a				
Protein Accession Number	Protein/Organism/Length	NOV80a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAM76703	Nuclear protein p30 - Homo sapiens (Human), 271 aa (fragment).	2..272 1..271	271/271 (100%) 271/271 (100%)	e-160
Q9CXS4	3110013H01Rik protein - Mus musculus (Mouse), 252 aa.	1..272 1..252	200/272 (73%) 213/272 (77%)	e-108
AAM29678	Hypothetical 14.6 kDa protein - Caenorhabditis elegans, 129 aa.	143..261 4..122	63/119 (52%) 77/119 (63%)	2e-32
Q9LFK7	Hypothetical 15.0 kDa protein - Arabidopsis thaliana (Mouse-ear cress), 135 aa.	142..272 5..135	61/132 (46%) 86/132 (64%)	2e-30
Q9AMY1	ID747 - Bradyrhizobium japonicum, 407 aa.	143..259 267..383	53/117 (45%) 75/117 (63%)	8e-28

5

Pfam analysis predicts that the NOV80a protein contains the domains shown in the Table 80E.

Table 80E. Domain Analysis of NOV80a			
Pfam Domain	NOV80a Match Region	Identities/ Similarities for the Matched Region	Expect Value

10

Example 81.

The NOV81 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 81A.

Table 81A. NOV81 Sequence Analysis			
	SEQ ID NO: 279	1051 bp	
NOV81a, CG56125-01 DNA Sequence	CCGCTGCTACCCGGCATGTGCGCGGAGGCCTCGGGCCCGGCTGCCGCCGCGCCCCGTCCCTG GAAGCCCCAAGCCCTCGGGTCTGAGCCTGGCCCCCGCGCTACGGTCTCAAGCCCGTGAAC CCGAACAGCAAATACGTGAAGCTGAACGTGGGCGGCTCGTTGCACTACACCACGCTGCGCACC CTCACGGGACAGGACACCATGTCTAAAGCCATGTTACAGCGGCGCGTGGAGGTGCTGACCGAT GCCGGAGGTTGGGTGCTGATTACCGGAGCGCCGTCACTTTGGTACAATCCTCAATTACCTG CGGGATGGGTCTGTGCCACTGCCGAGAGTACGAGAGAACTGGGGGAGCTGCTGGGCGAAGCA CGCTACTACCTGGTGCAGGGCCTGATTGAGGACTGCCAGCTGGCGCTGCAGCAAAAAGGGAG ACGCTGTCCCCGCTGTGCCTCATCCCCATGGTGACATCTCCCGGAGGAGCAGCAGCTCCTG GCCAGCACCTCCAAGCCGTGGTGAAGCTCCTGCACAACCGCAGTAACAACAAGTACTCCTAC ACCAGCACTTCAGATGACAACCTACTTAAGAACATCGAGCTGTTTCGACAAGCTGGCCCTGCGC TTCCACGGGCGGCTACTCTTCTCAAGGATGTCTGGGGACGAGATCTGCTGCTGGTCTTTT TACGGGCAGGGCCGCAAAATCGCCGAGGTGTGCTGCACCTCCATTGTCTATCGGAGAAG AAGCAGACCAAGGTGGAATTTCCAGAGGCCCGGATCTTCGAGGAGACCTGAACATCCTCATC TACGAGACTCCCGGGGCCAGACCCAGCCCTCTGGAGGCCACAGGGGAGCAGCTGGAGCT GGTGGGGCTGGCCGCGGGAGGATGAAGAAACCGAGAGCACCGTGTCCGAGGATCCATGTC CGGCGCCATATACCCACGACGAGCTCCTCATGGCCAACAAATTGTCTTCAAGGACTGACCT CTGACCCCTCCCCTGCCTTCTTCTTGGCTTGGGACCCAGTCCC		
	ORF Start: ATG at 16		ORF Stop: TGA at 1003
	SEQ ID NO: 280	329 aa	MW at 36357.0kD
NOV81a, CG56125-01 Protein Sequence	MSAEASGPAAAAAPSLKPKSPGLEPSPAAYGLKPLTPNSKYVKLVNVGSLHYTTLRLTLTGQD TMLKAMFSGRVEVLTDAGGWVIDRSGRHFGTILNLYLRDGSVPLPESTRELGELEARYLV QGLIEDCQLALQKRETLSPCLIPMVTSPREEQQLLASTSKPVVKLLHNRSNNKYSYTS DNLLKNIELFDKLALRFHGRLLFLKDLVLDGDEICWVSFYGQGRKIAEVCCTSIYATEKKQTKV EFPEARIFEETLNILYIETPRGPDALLEATGAAGAGGAGRGEDEENREHRVRRIHVRRHIT HDERPHGQQIVFKD		
	SEQ ID NO: 281	852 bp	
NOV81b, CG56125-02 DNA Sequence	TTTTCTTAGTTCTCAACTTAACCTTATTTCAATAATTTAATAGAAAATTAATAATAAAT AATATGAAACAGACTGATAACGCTGAGCTGGGCAGGCCAGGCCAGTCTAGTACAAAGTTAAG GAGGTAGGGAGGATGGTGGGGAGGAGGGGGCGGACTACCTGACAGGACGCGGGAGGCTGCTCA GACTGTGGTGATGTGAGGAAGGGCCGACACTTTGGCATGGACGATGCACTAAAAAAGAGAA AGGGAATTCTAAATCCCTCTTAACAGCTGGAGAGGGAAGGACGAGGGCCAGGGTGGGGACA AGTGTGGCTTCGGAAGGCTCTGAGTGGTGGGGCCGGAATGTACCATGTTGTAGCAATGGGG TTGGGACGGGTGGAGAAGGGCCAAAGTGAGCTGTGCCATGCAATGAAGGACAGAGGAGGACC CACGACTTGGCCAGCAGAGCCGGGGCAAAGTCTGGGAAGGGAGGGAAAGAGAGAGGGACTG GGTCCCAAGCAAGAGGAAGGCAGGGGGAGGGTCAGAGGTCAAGTCTTGAAGACAATTTGTTG GCCATGAGGACGCTCGTCGTGGGTGATATGGCGCCGGACATGGATCCTGCGGACACGGTGCTC TCGTTCTCTTCATCCTCCCGCGCCAGCCCAACAGCTCCAGCTGCTCCCTGTGGCTC CAGGAGGGCTGGGTCTGGGCCCCGGGAGTCTCGTAGATGAGGATGTTAGGGTCTCCTCGAA GATCCGGGCTCTGGAATTCACCTGCAAAAGGCCAGCCGGCCCCAGTCTCTCTTCTGGT GCAGTCACGAGGGCCATCCCCCTGCCTAGGNN		
	ORF Start: ATG at 361		ORF Stop: TAG at 847
	SEQ ID NO: 282	162 aa	MW at 17258.6kD
NOV81b, CG56125-02 Protein Sequence	MLLAMGLGRVEKGQSELCHAMKQRRTHDLASRAGAKVWEGEGKREGLGPKARGRQEGQRSV LEDNLLAMRTLIVGDMAPMDPADTVLSVLFILPAASPTSSSCSPCGLQEGWVWAPGSLVDED VQGLLEDPLWKFHLQKASRPQLLPSGAVTQGHPPA		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 81B.

Table 81B. Comparison of NOV81a against NOV81b.		
Protein Sequence	NOV81a Residues/ Match Residues	Identities/ Similarities for the Matched Region

NOV81b	126..132 127..133	6/7 (85%) 7/7 (99%)
--------	----------------------	------------------------

Further analysis of the NOV81a protein yielded the following properties shown in Table 81C.

Table 81C. Protein Sequence Properties NOV81a	
PSort analysis:	0.6500 probability located in cytoplasm; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV81a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 81D.

Table 81D. Geneseq Results for NOV81a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV81a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM39908	Human polypeptide SEQ ID NO 3053 - Homo sapiens, 329 aa. [WO200153312-A1, 26-JUL-2001]	1..329 1..329	329/329 (100%) 329/329 (100%)	0.0
ABB06073	Human NS protein sequence SEQ ID NO:165 - Homo sapiens, 323 aa. [WO200206315-A2, 24-JAN-2002]	34..324 34..323	213/293 (72%) 253/293 (85%)	e-123
AAB94285	Human protein sequence SEQ ID NO:14723 - Homo sapiens, 310 aa. [EP1074617-A2, 07-FEB-2001]	34..324 22..310	213/293 (72%) 252/293 (85%)	e-121
AAM94003	Human stomach cancer expressed polypeptide SEQ ID NO 76 - Homo sapiens, 310 aa. [WO200109317-A1, 08-FEB-2001]	34..324 22..310	213/293 (72%) 252/293 (85%)	e-121
AAB42647	Human ORFX ORF2411 polypeptide sequence SEQ ID NO:4822 - Homo sapiens, 195 aa. [WO200058473-A2, 05-OCT-2000]	72..252 1..181	181/181 (100%) 181/181 (100%)	e-102

10

In a BLAST search of public sequence databases, the NOV81a protein was found to have homology to the proteins shown in the BLASTP data in Table 81E.

Table 81E. Public BLASTP Results for NOV81a				
Protein Accession Number	Protein/Organism/Length	NOV81a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8WZ19	Hypothetical 36.4 kDa protein - Homo sapiens (Human), 329 aa.	1..329 1..329	329/329 (100%) 329/329 (100%)	0.0
Q96SA1	TNFAIP1-like protein - Homo sapiens (Human), 329 aa.	1..329 1..329	327/329 (99%) 328/329 (99%)	0.0
Q96P93	Polymerase delta-interacting protein 1 - Homo sapiens (Human), 329 aa.	1..329 1..329	321/329 (97%) 325/329 (98%)	0.0
Q96SU0	CDNA FLJ14637 fis, clone NT2RP2001327, moderately similar to tumor necrosis factor, alpha-induced protein 1 - Homo sapiens (Human), 310 aa.	34..324 22..310	213/293 (72%) 252/293 (85%)	e-121
Q9H3F6	MSTP028 - Homo sapiens (Human), 313 aa.	34..324 25..313	213/293 (72%) 252/293 (85%)	e-121

- 5 PFam analysis predicts that the NOV81a protein contains the domains shown in the Table 81F.

Table 81F. Domain Analysis of NOV81a			
Pfam Domain	NOV81a Match Region	Identities/ Similarities for the Matched Region	Expect Value
K_tetra	41..138	35/111 (32%) 71/111 (64%)	2.8e-27

Example 82.

- 10 The NOV82 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 82A.

Table 82A. NOV82 Sequence Analysis			
	SEQ ID NO: 283	630 bp	
NOV82a, CG57113-01 DNA	GTTCCGGGAGCCGCGGCTTATGGTGCAGACATGGCCAAGTCCAAGAACACACACACAACC AGTCCCGAAAATGGCACAGAAATGGTATCAAGAAACCCGATCACAAAGATACGAATCTCTTA		

Sequence	AGGGGGTGGACCCCAAGTTCCTGAGGAACATGCGCTTTGCCAAGAAGCACAACAAAAGGGCC TAAAGAAGATGCAGGCCAACATGCCAAGGCCATGAGTGACGTGCCGAGGCTATCAAGGGCC TCGTAAAGCCCAAGGAGGTTAAGCCCAAGATCCCAAAGGGTGTGAGCCGAAGCTCGATCGAC TTGCCTACATTGCCCCACCCCAAGCTTGGGAAGCGTGCTCGTGCCCGTATTGCCAAGGGGCTCA GGCTGTGCCGCCAAAGGCCAAGGCCAAGGCCAAGGCCAAGGATCAAACCAAGGCCCAGGCTG CAGCCCCAGCTTCAGTTCCAGCTCAGGCTCCCAAACGTACCCAGGCCCTACAAAAGGCTTCAG AGTAGATATCTCTGCCAACATGAGGACAGAAGGACTGGTGCGACCCCCACCCCGCCCTGG GCTACCATCTGCATGGGGCTGGGGTCCTCTGTGCTACTGGTACAAATAAACCTGAGGCAGGA		
	ORF Start: ATG at 30		ORF Stop: TAG at 507
	SEQ ID NO: 284	159 aa	MW at 17751.9kD
NOV82a, CG57113-01 Protein Sequence	MAKSKNHTTHNQSRKWHRNGIKKPRSORYESLKGVDPKFLRNMRFKKHNKGLKKMQANNAK AMSARAEAIKALVKPKVKPKIPKGVSRKLDRLAYIAHPKLGKRARARIAKGLRLCRPKAKAK AKAKDQTKAQAAPASVPAQAPKRTQAPT KASE		
	SEQ ID NO: 285	600 bp	
NOV82b, CG57113-03 DNA Sequence	CGCGGCTTATGGTGCAGACATGGCCAAGTCCAAGAACCACACCACACCAACAGTCCCGAAA ATGGCACAGAAATGGTATCAAGAAACCCCGATCACAAGATACGAATCTCTTAAGGGGGTGA CCCCAAGTTCCTGAGGAACATGCGCTTTGCCAAGAAGCACAACAAAAGGGCTTAAAGAAGAT GCAGGCCAACAAATGCCAAGGCCATGAGTGACGTGCCGAGGCTATCAAGGCCCTCGTAAAGCC CAAGGAGGTTAAGCCCAAGATCCCAAAGGGTGTGAGCCGCAAGCTCGATCGACTTGCTTACAT TGCCACCCCAAGCTTGGGAAGCGTGCTCGTGCCCGTATTGCCAAGGGGCTCAGGCTGTGCCG GCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGCCAAGGATCAAACCAAGGCCCAGGCTGCAGC CCCAGCTTCAGTTCCAGCTCAGGCTCCCAAACGTACCCAGGCCCTTACAAAGGCTTCAGAGTA GATATCTCTGCCAACATGAGGACAGAAGGACTGGTGCGACCCCCACCCCGCCCTGGGCTA CCATCTGCATGGGGCTGGGGTCCTCTGTGCTA		
	ORF Start: ATG at 20		ORF Stop: TAG at 503
	SEQ ID NO: 286	161 aa	MW at 17951.1kD
NOV82b, CG57113-03 Protein Sequence	MAKSKNHTTHNQSRKWHRNGIKKPRSORYESLKGVDPKFLRNMRFKKHNKGLKKMQANNAK AMSARAEAIKALVKPKVKPKIPKGVSRKLDRLAYIAHPKLGKRARARIAKGLRLCRPKAKAK AKAKDQTKAQAAPASVPAQAPKRTQAPT KASE		
	SEQ ID NO: 287	579 bp	
NOV82c, CG57113-02 DNA Sequence	ACTCACTATAGGGCTCGAGCGCGCTTCGGGAGCCGCGGCTTATGGTGCAGACATGGCCAAGT CCAAGAACCACACCACACCAACAGTCCCGAAAATGGCACAGAAATGGTATCAAGAAACCC GATCACAAGATACGAATCTCTTAAGGGGGTGGACCCCAAGTTCCTGAGGAACATGCGCTTTG CCAAGAAGCACAACAAAAGGGCTTAAAGAAGATGCAGGCCAACAAATGCCAAGGCCATGAGTG CACGTGCCGAGGCTATCAAGGCCCTCGTAAAGCCCAAGGAGGTTAAGCCCAAGATCCCAAAGG GTGTCAGCCGCAAGCTCGATCGACTTGCTTACATTGCCACCCCAAGCTTGGGAAGCGTGCTC TGCCCGTATTGCCAAGGGGCTCAGGCTGTGCCGGCCAAAGGCCAAGGCCAAGGCCAAGGCCA AGGCCAAGGATCAAACCAAGGCCCAGGCTGCAGCCCCAGCTTCAGTTCCAGCTCAGGCTCCCA AACGTACCCAGGCCCTTACAAAGGCTTCAGAGTAGATATCTCTGCCAACATGAGGACAGAAGG GCCTGGTGTCCA		
	ORF Start: ATG at 54		ORF Stop: TAG at 537
	SEQ ID NO: 288	161 aa	MW at 17951.1kD
NOV82c, CG57113-02 Protein Sequence	MAKSKNHTTHNQSRKWHRNGIKKPRSORYESLKGVDPKFLRNMRFKKHNKGLKKMQANNAK AMSARAEAIKALVKPKVKPKIPKGVSRKLDRLAYIAHPKLGKRARARIAKGLRLCRPKAKAK AKAKDQTKAQAAPASVPAQAPKRTQAPT KASE		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 82B.

Table 82B. Comparison of NOV82a against NOV82b and NOV82c.		
Protein Sequence	NOV82a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV82b	1..159 1..161	116/161 (72%) 116/161 (72%)
NOV82c	1..159 1..161	116/161 (72%) 116/161 (72%)

Further analysis of the NOV82a protein yielded the following properties shown in Table 82C.

Table 82C. Protein Sequence Properties NOV82a	
PSort analysis:	0.9840 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV82a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 82D.

Table 82D. Geneseq Results for NOV82a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV82a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW71709	Heparan sulfate/heparin interacting protein - Homo sapiens, 159 aa. [WO9838214-A1, 03-SEP-1998]	1..159 1..159	159/159 (100%) 159/159 (100%)	4e-87
AAM80294	Human protein SEQ ID NO 3946 - Homo sapiens, 167 aa. [WO200157190-A2, 09-AUG-2001]	1..159 1..167	159/167 (95%) 159/167 (95%)	6e-85
ABG26996	Novel human diagnostic protein #26987 - Homo sapiens, 223 aa. [WO200175067-A2, 11-OCT-2001]	1..149 12..156	144/149 (96%) 144/149 (96%)	6e-77
ABG11437	Novel human diagnostic protein #11428 - Homo sapiens, 189 aa. [WO200175067-A2, 11-OCT-2001]	5..159 32..189	146/158 (92%) 149/158 (93%)	2e-76

ABG07644	Novel human diagnostic protein #7635 - Homo sapiens, 571 aa. [WO200175067-A2, 11-OCT-2001]	1..154 1..150	139/154 (90%) 141/154 (91%)	1e-72
----------	--	------------------	--------------------------------	-------

In a BLAST search of public sequence databases, the NOV82a protein was found to have homology to the proteins shown in the BLASTP data in Table 82E.

Table 82E. Public BLASTP Results for NOV82a				
Protein Accession Number	Protein/Organism/Length	NOV82a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH08926	Ribosomal protein L29 - Homo sapiens (Human), 159 aa.	1..159 1..159	159/159 (100%) 159/159 (100%)	1e-86
P47914	60S ribosomal protein L29 (Cell surface heparin binding protein HIP) - Homo sapiens (Human), 158 aa.	2..159 1..158	158/158 (100%) 158/158 (100%)	4e-86
Q9XS36	Ribosomal protein L29/heparin/heparan sulfate interacting protein - Sus scrofa (Pig), 160 aa.	1..159 1..160	137/160 (85%) 143/160 (88%)	4e-72
R6RT43	ribosomal protein L29, cytosolic [validated] - rat, 156 aa.	1..157 1..155	127/157 (80%) 137/157 (86%)	3e-67
P25886	60S ribosomal protein L29 (P23) - Rattus norvegicus (Rat), 155 aa.	2..157 1..154	126/156 (80%) 136/156 (86%)	1e-66

5

Pfam analysis predicts that the NOV82a protein contains the domains shown in the Table 82F.

Table 82F. Domain Analysis of NOV82a			
Pfam Domain	NOV82a Match Region	Identities/ Similarities for the Matched Region	Expect Value
Ribosomal_L29e	3..42	31/40 (78%) 40/40 (100%)	2.3e-19

10

Example 83.

The NOV83 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 83A.

Table 83A. NOV83 Sequence Analysis

	SEQ ID NO: 289	1409 bp	
NOV83a, CG59536-01 DNA Sequence	GGCACAGCCAGGAACGTTGCTTTGGAGAATCCTGCAGATAAGGCTTTTCCAAAAGCGCGAG CATCTTGTGTATTACAGATACCCATCGTCGTCAGTCATGGCTAGCATCACTGCGTGTGTGGG TAACAGCAGGCAGCAGAAATGCACCTTTGCCGCCTTGGGCCATTCCATGTTGAGGTCTCTGGG GAGGAGTCTCTGTCTTTAGTGGTCAAAATGGCAGAGAGAAACATGAAGTTGTTCTCAGGAAG AGTGGTGCCAGCCAGGGGAAAGAAACCTTTGAAAACCTGGCTGATCCAAGTCAATGAGGTCTCT GCCAGATTGGAGTATGTCTGAGGAGGAAAACTCAAGCGCTTGATGAAAACACTTAGGGGCC TGCCCGGAGGTATGCGTTTCTCAGGCGGCCAACCCCACTTAAGTGTAGCAGATTTCTT GCGGGCAATGAAATTGGTGTGTTGGGAGTCTGAAAGCAGTGTGACTGCCCATGGTAAATTTT TAACACCTGCAGGCACAAGGGGAGAAAGCCTCCCTTTATGTGATCCGTTTAGAGGTGCAGCT CCAGAATGCTATTAGGCAGGCATCCTAGCTGAGAAAGATGCAAACCAGACTCGCTTGAACA GCTTCTTTAGGCGCTGAGCTGAATAGGGACCTGCGCTTACGGCTTAAGCATCTTCTCAGGAT GTATGCAAATAAGCAGGAGCGGCTTCCCAATTTCTGGAGTTAATCAAGATGATAAGGGAGGA AGAGGATTGGGATGATGCTTTTATTAAACGGAAGCGGCCGAAAAGGTCTGAGCCAATAATGGA GAGGGCAGCCAGCCCTGTGGCATTTCAGGGCGCCAGCCAATAGCAATCAGCAGTGTGACTG TAAGTCAACGTGATAGAAATAGATGATACCTTGATGACTCTGATGAGGATGTGATCCTGCT GGTGTCTCTGTACCTTCACTGACACCTACAGGTGCCCCCTCCCTTCAGAGGAAGGCCAGACC TCTGGATCAAGTGTGTTATTGATCCCCCAACAATTCTGGGGCTCAGTCTCTTTCTACCAG TGGTGGTCTGGGTATAAGAAATGATGGTCTGGGAATATTCTGATAGCCAGGAAGCGAAAATA CACAACCCGCTGTTCATATTGTGGGGAGGAGGGCCACTCAAAGAAACCTGTGACAATGAGAG CAACAAGGCCAGGTTTTTGAGAACTGTATCATCACCTGCAGGAGCTGACACATACAGAGGA GAGGTCAAAGAGGTCCCTGGAGAACACAGTGTGCTTCTGAGCCACAGTAAGGATCTAGTCC AGCCCTAAATGAGTCTTACTGTATTAGAGTCTGGTAATGGGAATAACAGGAGAGGGGGGT GGGTTTCTAACTGCATGAATTA		
	ORF Start: ATG at 101		ORF Stop: TAA at 1310
	SEQ ID NO: 290	403 aa	MW at 45159.8kD
NOV83a, CG59536-01 Protein Sequence	MASITACVGNRQONAPLPWAHSMRLSLGRSLPLVVKMAERNMKLFSGRVVPAQKETFEN WLIQVNEVLPDWSMSEEEKLRLMKTLRGPAREVMRLQANPNLSVADFLRAMKLVFGESES SVTAHGKFFNTLQAQGEKASLYVIRLEVQLQNAIQAGILAEKDANQTRLQQLLGAELNRDLR FRLKHLRLMYANKQERLPNLELTKMIREEDWDDAFIKRRPKRSEPIMERAAASPVAFQGAQ PIAISSADCNCNVEIDDTLDDSDVDVILVVSPLYSLTPTGAPPFRGRARPLDQVLVIDSPNN SGAQLSTSGGSGYKNDGPGNIRARKRKYTTRCSYCGEEHGSKETCDNESNKAQVFENLIIT LQELTHTEERSKEVPGEHSDASEPQ		
	SEQ ID NO: 291	1360 bp	
NOV83b, CG59536-02 DNA Sequence	GTTGCTTTGGAGAATCCTGCAGATAAGGCTTTTCCAAAAGCGCGAGCATCTTGTGTATTCA GATACCTTACCGTCGTCAGTCATGGCTAGCATCACTGCGCGTGTGGGTAAACAGCAGGCAGCAG AATGCACCTTTGCGCCTTGGGCCATTCCATGTTGAGGTCTCTGGGGAGGATCTCTGTCTCT TTAGTGGTCAAAATGGCAGAGAGAAACATGAAGTTGTTCTCAGGAAGAGTGGTGCCAGCCCAG GGGAAAGAAACCTTTGAAAACCTGGCTGATCCAAGTCAATGAGGTCTGCCAGATTGGAGTATG TCTGAGGAGGAAAACTCAAGCGCTTGATGAAAACACTTAGGGGCCCTGCCCGGAGGTCTATG CGTTTGTCTCAGGCGGCCAACCCCACTAAGTGTAGCAGATTCTTTCGGGCAATGAAATTG GTGTTTGGGGAGTCTGAAAGCAGTGTGACTGCCCATGGTAAATTTTAAACACCTGCAGGCA CAAGGGGAGAAAGCCTCCCTTTATGTGATCCGTTTAGAGGTGCAGCTCCAGAATGCTATTAG GCAGGCATCTAGCTGAGAAAGGTGCAACCAGACTCGCTTGCAACAGCTCTTTTAGGCGCT GAGCTGAATAGGACCTGCGCTTACGGCTTAAGCATCTTCTCAGGATGTATGCAAATAAGCAG GAGCGCTTCCCAATTTCTGGAGTTAATCAAGATGATAAGGGAGGAAGAGGATTGGGATGAT GCTTTTATTAAACGGAAGCGGCCGAAAAGGTCTGAGCCAATAATGGAGAGGGCAGCCAGCCCT GTGGCATTTCAGGGCGCCAGCCAATAGCAATCAGCAGTGTGACTGTAAGTCAACGTGATA GAAATAGATGATACCTTGTGACTCTGATGAGGATGTGATCCTGGTGGTGTCTCTGTACCTT TCACTGACACCTACAGTGCCCTCCCTTCAGAGGAAGAGCCAGACCTCTGGATCAAGTGTG GTTATTGATTCCCCCAACAATTCGGGGCTCAGTCTCTTCTACAGTGGTGGTCTCTGGGTAT AAGAATGATGGTCTGGGAATATTCTGATAGCCAGGAAGCGAAAATACACAACCCGCTGTTCA TATTGTGGGGAGGAGGCCACTCAAAGAAACCTGTGACAATGAGAGCAACAAGGCCAGGTT TTTGAGAACTGATCATCACCTGCAGGAGCTGACACATACAGAGGAGAGGTCAAAGAGGTC CCTGGAGAACACAGTGTGCTTCTGAGCCACAGTAAGGATCTAGTCCAGCCCTAAATGAGTCC TTGACTGTATTAGAGTCTGGTAATGGGAATAACAGG		
	ORF Start: ATG at 85		ORF Stop: TAA at 1294
	SEQ ID NO: 292	403 aa	MW at 45154.9kD
NOV83b, CG59536-02 Protein Sequence	MASITARVGNRQONAPLPWAHSMRLSLGRSLPLVVKMAERNMKLFSGRVVPAQKETFEN WLIQVNEVLPDWSMSEEEKLRLMKTLRGPAREVMRLQANPNLSVADFLRAMKLVFGESES SVTAHGKFFNTLQAQGEKASLYVIRLEVQLQNAIQAGILAEKGANQTRLQQLLGAELNRDLR		

Sequence	FRLKHLRLMYANKQERLPNFLELIKMIREEEDWDDAFIKRRPKRSEPIMERAASPVAFQGAQ PIAISSADCNCNVIEIDDTLDDSDVDILVVSLYPSLTPTGAPPRGRARPLDQVLVIDSPNN SGAQLSTSGGSGYKNDGPGNIRRARKRKYTTSCSYCGEEGHSKETCDNESNKAQVFENLIIT LQELTHTEERSKEVPGEHSDASEPQ
----------	--

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 83B.

Table 83B. Comparison of NOV83a against NOV83b.		
Protein Sequence	NOV83a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV83b	1..403	384/403 (95%)
	1..403	384/403 (95%)

Further analysis of the NOV83a protein yielded the following properties shown in Table 83C.

Table 83C. Protein Sequence Properties NOV83a	
PSort analysis:	0.7000 probability located in nucleus; 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen); 0.0000 probability located in endoplasmic reticulum (membrane)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV83a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 83D.

Table 83D. Geneseq Results for NOV83a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV83a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM51624	KIAA0883-44 polypeptide - Unidentified, 403 aa. [WO200183540-A1, 08-NOV-2001]	1..403 1..403	401/403 (99%) 401/403 (99%)	0.0
AAB60478	Human cell cycle and proliferation protein CCYPR-26, SEQ ID NO:26 - Homo sapiens, 402 aa. [WO200107471-A2, 01-FEB-2001]	1..402 1..401	339/403 (84%) 364/403 (90%)	0.0
AAM25693				e-154

	NO:1208 - Homo sapiens, 337 aa. [WO200153455-A2, 26-JUL-2001]	1..337	298/339 (87%)	
AAB12528	Human Ma4 protein SEQ ID NO:11 - Homo sapiens, 283 aa. [JP2000146982- A, 26-MAY-2000]	22..226 62..260	84/205 (40%) 132/205 (63%)	7e-37
AAB74701	Human membrane associated protein MEMAP-7 - Homo sapiens, 353 aa. [WO200112662-A2, 22-FEB-2001]	4..220 119..332	87/218 (39%) 127/218 (57%)	3e-36

In a BLAST search of public sequence databases, the NOV83a protein was found to have homology to the proteins shown in the BLASTP data in Table 83E.

Table 83E. Public BLASTP Results for NOV83a				
Protein Accession Number	Protein/Organism/Length	NOV83a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q8TE36	CU46H11.1 (novel protein) - Homo sapiens (Human), 403 aa.	1..403 1..403	403/403 (100%) 403/403 (100%)	0.0
AAH31241	Similar to cU46H11.1 (novel protein) - Homo sapiens (Human), 402 aa.	1..402 1..401	339/403 (84%) 364/403 (90%)	0.0
Q9CZA5	2810028A01Rik protein - Mus musculus (Mouse), 402 aa.	1..403 1..402	291/404 (72%) 332/404 (82%)	e-161
Q8VD24	RIKEN cDNA 1500031H04 gene - Mus musculus (Mouse), 393 aa.	1..393 1..392	268/394 (68%) 322/394 (81%)	e-151
Q9DB17	1500031H04Rik protein - Mus musculus (Mouse), 393 aa.	1..393 1..392	267/394 (67%) 322/394 (80%)	e-150

5

PFam analysis predicts that the NOV83a protein contains the domains shown in the Table 83F.

Table 83F. Domain Analysis of NOV83a			
Pfam Domain	NOV83a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-CCHC	347..364	6/18 (33%) 12/18 (67%)	0.059

Example 84.

The NOV84 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 84A.

Table 84A. NOV84 Sequence Analysis			
	SEQ ID NO: 293	3061 bp	
NOV84a, CG59794-01 DNA Sequence	GTTTGAAGTAGATCTTTGGTGAGCACCCTCACATCTCTAAATGTGGGTTTCAGACTCTGAATT TTTGCCACTTTCTTATTGCTTCTCTTGACAGGGATCATGGCCAGGTAGCAGTGTCCACCCCT GCCTGTTGAAGAAGAGTCTCTCAGAGACCAGGATGGTGGTGACATTCTCTGCTGTGCCCCCT CGAATCCATGTGTAAAGAACTGGCCAAAGTCCAAGGCAGAAGTGGCCTGCATCGCAGTGTACGA AACAGACGTGTTGTGCTCGGAACCGAGAGAGGATGCGCTTTTGTAAATGCCAGGACGGATTT TCAGAAAGATTTTGCAAATACTGTAGGGCAGAGGGACTGTGTGAGGTGAAACCTCCCTGCCC TGTGAACGGGATGCAGGTCCACTCGGGCGAAACGGAATACTCAGGAAGGCAGTGGAGGACTA TTTCTGCTTTTGTATCAGAAAGCCTTAGGGACAACAGTGATGGTGCCTGTTCCTATGAGAA GATGCTGCGAGACCAGTCCGCTGTGGTAGTGACGGGGCTTCCGGAAGCGTTCCTTTCAACA CCCTGAGAATTACGACCTTGCAACCCTGAAATGGATTTTGGAGAACAAAGCAGGGATTTCAAT CATCATAAATAGGAGACCCTTCCTAGGACCAGAGAGTCAAGTGGGTGGCCCTGGGATGGTAAC AGATGCGGAGAGATCCATAGTATCACCAGTGAAAGCTGCGGCCCATCAATGTGAAAACTGA ACCCATGGAAGATTCTGGCATTCTCACTGAAAGCAGAAGCTGTCTCAGTCAAGAAAGAAATCAGA AGATCCTAATTACTATCAATATAATATGCAAGGTAATAGCCACCCTTCTTCCACAAGCAATGA AGTAATAGAAATGGAATTACCAATGGAACAAGATCCACTCCGCTGGTCCCTTCAGAGTGAACC AAATGAGGACCCTGAAGCCGAGGTGAAATCGAAGGTGGAAACACAAATTCATCCAGTGTAC AAATTCTGCAGCAGGTGTTGAAGATCTTAACATCGTTCAAGTGACTGTTGATAATGAGAAGGA AAGATTATCAAGCATTGAAAAGATTAAACAGCTAAGAGAACAAAGTAAATGACCTCTTTAGCCG AAAATTTGGTAAAGCAATTGGCGTGGATTTCCCTGTGAAAGTTCCTACAGGAAGATCACATT CAACCTCGGCTGTGTGGTGATTGATGGCATGCCCCCGGGGGTGGTATTCAAGGCCCCCGGCTA TCTGGAATCAGTTCATGAGGAGGATCTTGAGGCGAGCTGAGTTTATCAAAATTCACAGTCAT CAGGCCGCTTCCAGGGCTTGAGCTCAGTAATGTGGGAAAACGCAAGATAGACCAGGAGGGCGG TGTGTTTCAAGAAAAGTGGGAGAGAGCGTATTTCTTCGTGGAAGTACAGAATATTCCAACATG TCTCATATGCAACAAAGCATGTCTGTGTCCAAAGAATATAACCTAAGACGCCACTATCAAAC CAATCACAGCAAGCATTATGACCAGTATATGGAAGAATGCGTGACGAGAAGCTTCACGAGCT GAAAAAAGGGCTCAGGAAGATCTCTTAGGCTTGTGACACCCAGTGTCCCGAGCAAAAAACA AGTGTGTGCAACCCAAAGTCCAACCCAGAAATCCCCGTCAGCCTGTAGAGGACCTAGCTGG GAACTTATGGGAGAAGTTACGTGAAAAAATCAGGTCTTTTGTGGCATATTCTATCGCAATCGA TGAGATCACGGATATAAATAATACCACCCAGTTGGCCATATTATCCGCTGGTGTCCATGAGAA TTTCGATGTGTCCGAAGAACTTCTGGACACGGTGCCCATGACGGGTACAAATCTGGCAACGA GATCTTTTCGCGTGTGTGAGAAGAGCCTGAAAAAGTTCTGTATCGACTGGTCAAAATAGTAGA CGTGGCCTCCACTGGCACCCAGCGATGGTGGATGCCAATAACGGGCTTGTACAAAACTGAA GTCCAGGGTGGCGACGTTCTGCAAGGGTGGGAACTGAAGTCCATCTGTTGATAATTATCATCC GGAATCACTCTGTGCTCAGAAGTTGAAGATGGACACGTCATGGACGTGGTAGTGAAGTCCGT GAACTGGATATGCTCCCGGGGACTGAACACAGTGAGTTCAACCTTGCTCTATGAGCTGGGA CAGCCAGTATGGTAGCCTCCTGTACTACACGGAGATTAAGTGGCTCAGTCGCGGGCTCGTGCT AAAGAGATTTTTCGAATCCTTGAAGAAATCGACTCCTTATGTATCCAGAGGGAACCCCT GCCTCAACTGAGCTCCATAGATTGGATCCGAGACCTGGCCTTCTTGGTTGACATGACGATGCA TCTGAACGCTTTGAACATCTCTCTCAAGGACACTCCCAAATCGTCACGCAGATGTATGACCT GATCCGGGCGTTCCTAGCAAAATCTGCTCTGCGGAGACTCATTGACGAGGAATAATCTGGC CCACTTTCCACCCCTGAAATTTGGCTTCCAGAAATGAAAGCGATGGCCTGAACTACATTCCCAA AATCGCGGAATCAAGACCGAATTCCAGAAAAGGCTGTCTGATTTCAAACTCTACGAAAGCGA ACTGACTCTGTTCAAGTCCCGGTTCTCCACGAAGATCGACAGTGTGCACGAGGAGCTCCAGAT GGAGGTTATCGACTGCAATGCAACACGGTCTGAAGACGAAATACGACAGGTGGGAATACC AGAATTCTACAAGTACCTCTGGGGTAGCTACCCGAAATACAAGCACCATTGCGCAAGATTCT TTCCATGTTTCGGGAGCACCTACATTGCGAACAGCTGTTCTCATTATGAACTGAGCAAAAC AAAATACTGCTCCAGTTAAAGGATTCCAGTGGGATTCTGTACTCCACATCGCAACGTGATG GAGAGAAAACCTCTGGCAGGGCCCTATGGTGGGAAAGGCTGGAGTCTTCTAGTCCCAAGGGAT TGGGAGATGACAAATGAATTTTCTTTCTTTTGA		
	ORF Start: ATG at 41		ORF Stop: TGA at 2957
	SEQ ID NO: 294	972 aa	MW at 109960.0kD
NOV84a, CG59794-01 Protein	MWVSDSEFLPLSYCFSTGIMAVAVSTLPVEEESSETRMVVTVFLVSALESMCKELAKSKAE VACIAVYETDVFVVGTERGCATFVNARTDFQKDFAKYCRAEGLCEVKPPCPVNGMVHSGTEI LRKAVEDYFCFCYQKALGTTVMVPVPEKMLRDQSAVVVQGLPEGVAFQHPENYDLATLKWIL		

Sequence	ENKAGISFIINRRPFLGPESQLGGPGMVTDAERSIVSPSESCGPINVKTEPMEDSGISLKAEA VSVKKESEDPNYYQYNMQGNSHPSSTSNEVIEMELPMEQDSTPLVPSEEPNEDPEAEVKIEGG NTNSSSVTNSAAGVEDLNIQVTVDNEKERLSSIEKIKQLREQVNDLFSRKFKAIGVDFPVK VPYRKITFNPFCVVIDGMPGVVFKAPGYLEISSMRRIEAAEFIKFTVIRPLPGLELSNVGK RKIDQEGRVFQEKWERAYFFVEVQNIPTCLICKQSMSVSKEYNLRRHYQTNHSHKYDQYMERM RDEKLHELKGLRKYLLGLSDTECEPQKQVFANPSPTQKSPVQPVEDLAGNLWEKLREKIRSF VAYSIAIDEITDINNTTQLAIFIRGVDFVSEELLDTVPMTGKSGNEIFSRVEKSLKKFC IDWSKLVSVASTGTPAMVDANGLVTKLSRVATFCKGAELKSICCIHPESLCAQKLKMDHV MDVVVKSVMWICSRGLNHSEFTTLLYELDSQYGSLLYYTEIKWLSRGLVLKRFESLEIDSF MSSRGKPLPQLSSIDWIRDLAFLVDMTMLNALNISLQGHISQIVTQMYDLIRAFKLCLWET HLTRNNLAHFPTLKLASRNESDGLNYPKIAELKTEFQKRLSDFKLYESELTFSSPFSTKID SVHEELQMEVIDLQCNVTLKTKYDKVGIPEFYKYLWGSYPKYKHKHCAKILSMFGSTYICEQLF SIMKLSKTKYCSQLKDSQWDSVLHIAT		
	SEQ ID NO: 295	2031 bp	
NOV84b, CG59794-02 DNA Sequence	CAGACCTGCGGCGACAGAGCAAACTCCGTCTCAACAACAACGACAACAAAATTCAGTCTTC AGGTTTTCTTTAGAAAACCTGAAGATCTGGCCACAGCTGGCATCCTGGCAGCGTTGCTGGA GTTGAGGGTCAGCCGCTCCCTCTGACGGGTGGGTACCCCTCCTGTTAAACACGCGCTGCCCCGC CCCGCTTCTCCTCTCTGTCGCTCATCAAGCATTGCTGTTGTTTCTCATAGTAGTGATAA GAGAAAAGTGAAATATCTTTGTCTCCTGTCTCTGTCAAAAGTGGGAAAACGCAAGATAGACC AGGAGGGCCGTGTGTTTCAAGAAAAGTGGGAGAGAGCGTATTCTTCGTGGAAGTACAGAATA TTCCAACATGTCTCATATGCAAAACAAAGCATGTCTGTGTCCAAAGAAATAAATCAAGACGCC ACTATCAAAACCAATCACAGCAAGCATTATGACCAGTATATGGAAAGATGCGTGACGAGAAGC TTCACGAGCTGAAAAAAGGGCTCAGGAAGTATCTCTTAGGCTGTGACAGACCGAGTGTCCCG AGCAAAAACAAGTGTTTGCACAAACCAAGTCAACCCAGAAATCCCCCGTGACGCTGTAGAGG ACCTAGCTGGGAACCTATGGGAGAAGTTACGTGAAAAAATCAGGTCTTTGTGGCATATTCTA TCGCAATCGATGAGATCACGGATATAAATAATACCAACCCAGTTGGCCATATTCATCCGTGGTG TCGATGAGAATTTGATGTGTCCGAAGAATCTCTGGACACGGTGCCATGACGGGTACAAAAT CTGGCAACGAGATCTTTTCGCGTGTGAGAAGAGCCTGAAAAAGTTCTGTATCGACTGGTCTGA AATTAGTAAGCGTGGCCTCACTGGCACCCAGCGATGGTGGATGCCAATAACGGGCTGTCA CAAAACCTGAAGTCCAGGGTGGCGACGTTCTGCAAGGGTGGCAACTGAAGTCCATCTGTGTGA TAATTCTATCCGGAATCACTCTGTGCTCAGAAGTTGAAGATGGACCACGTGATGGACGTGGTAG TGAAGTCCGTGAAGTGGATATGCTCCCGGGGACTGAACACAGTGAGTTCACAACTTGCTCT ATGAGCTGGACAGCCAGTATGGTAGCCTCCTGTACTACACGGAGATTAAGTGGCTCAGTCGCG GGCTCGTGCTAAAGAGATTTTTCGAATCCTTGGAAGAAATCGACTCCTTCATGTCATCCAGAG GGAAACCCCTGCCTCAACTGAGCTCCATAGATTGGATCCGAGACCTGGCCTTCTTGTTGACA TGACGATGCATCTGAACGCTTTGAACATCTCTCTCAAGGACACTCCAAATCGTCACGCAGA TGTATGACCTGATCCGGGCGTCTCTAGCAAACTGTGCCTCTGGGAGACTCATTGACGAGGA ATAATCTGGCCCAGTTTCCACCCCTGAAATTTGGCTTCCAGAAATGAAGCGATGGCCTGAAT ACATTTCCAAAATCGCGGAACCTCAAGACCGAATTCAGAAAAGGCTGTCTGATTTCAAATCT ACGAAAGCGAACTGACTCTGTTCAGCTCCCGTCTCCACGAAGATCGACAGTGTGCACGAGG AGCTCCAGATGGAGGTTATCGACCTGCAATGCAACACGGTCTGAAGACGAAATACGACAAGG TGGGAATACCAAGATTTTACAAGTACCTCTGGGGTAGCTACCCGAAATACAAGCACCATTGCG CAAAGATTTCTTCCATGTTCTGGGAGCACCTACATCTGCGAACAGCTGTTCTCATATGCAAT TGAGCAAAACAAAATACTGCTCCAGTTAAAGGATTCCAGTGGGATTCTGTACTCCACATCG CAACGTGATGGAGAGAAAACCTCTGGCAGGGCCCTATGGTGGGAAAGGCTGGAGTCTTCTAGT CCCAAGGGATTGGGAGATGACAAAATGAATTTTTTTTCTTTTGTAGATGGAGTCTTGCTCT GTCGCCGCGAGGTCTG		
	ORF Start: ATG at 408		ORF Stop: TGA at 1896
	SEQ ID NO: 296	496 aa	MW at 57221.3kD
NOV84b, CG59794-02 Protein Sequence	MSVSKEYNLRRHYQTNHSHKYDQYMERMRDEKLHELKGLRKYLLGLSDTECEPQKQVFANPS PTQKSPVQPVEDLAGNLWEKLREKIRSFVAYSIAIDEITDINNTTQLAIFIRGVDFVSE LLDTVPMGTGKSGNEIFSRVEKSLKKFCIDWSKLVSVASTGTPAMVDANGLVTKLSRVATF CKGAELKSICCIHPESLCAQKLKMDHVMVVVKSVMWICSRGLNHSEFTTLLYELDSQYGS LYYTEIKWLSRGLVLKRFESLEIDSFMSRGKPLPQLSSIDWIRDLAFLVDMTMLNALNI SLQGHISQIVTQMYDLIRAFKLCLWETHLTRNNLAHFPTLKLASRNESDGLNYPKIAELKT EFQKRLSDFKLYESELTFSSPFSTKIDSVHEELQMEVIDLQCNVTLKTKYDKVGIPEFYKYL WGSYPKYKHKHCAKILSMFGSTYICEQLFSIMKLSKTKYCSQLKDSQWDSVLHIAT		
	SEQ ID NO: 297	3123 bp	
NOV84c, CG59794-03 DNA Sequence	CGAGTGGCGAGCAGGGGCTCGGCCGCCACACAGCCCGAAGCGTGCTCGTCCCCCGCGC GGGGCTCCCGGCGCGCGCTCGGCCATCGGCTGCTCCCGGTGGCCAGGCTCGGACTCCG CGGCCGCGCGCGCGCGCCAGCGCCCTCAGGGATCATGGCCAGGTAGCAGTGTCCACCCTG CCTGTTGAAGAAGAGTCTCTCTCAGAGACCAGGATGGTGGTGACATTCCTCGTGTCTGCCCTC GAATCCATGTGTAAAGAACTGGCCAAGTCCAAGGCAGAAGTGGCTGCATCGCAGGTACGAA		

	<p>ACAGACGTGTTTGTGCTCGGAACCGAGAGAGGATGCGCTTTTGTAAATGCCAGGACGGATTTT CAGAAAGATTTTGC AAAA TACTGCGTTGCAGAGGGACTGTGTGAGGTGAAACCTCCCTGCCCT GTGAACGGGATGCAGGTCCACTCGGGCGAAACGGAATACTCAGGAAGGCAGTGGAGGACTAT TTCTGCTTTTGTATCAGAAAGCCTTAGGGACAACAGTGTGGTCCCTGTTCCCTATGAGAAG ATGCTGCGAGACCAGTCCGCTGTGGTAGTGCAGGGGCTTCCGGAAGGCGTTGCCCTTCAACAC CCTGAGAATTACGACCTTGCAACCTGAAATGGATTTTGGAGAACAAGCAGGGATTTCATT ATCATAAATAGGAGACCTTCTAGGACCAGAGAGTCAGCTGGGTGGCCCTGGGATGGTAACA GATGCGGAGAGATCCATAGTATCACCAAGTGAAAGCTGCGGCCCCATCAATGTGAAAAGTGA CCCATGGAAGATTCTGGCATTTCAGTGAAGCAGAAAGCTGTCTCAGTCAAGAAAGAATCAGAA GATCCTAATTACTATCAATATAATATGCAAGGTAATAGCCACCCTTCTTCCACAAGCAATGAA GTAATAGAAATGGAATTACCAATGGAACAAGATCCACTCCGCTGGTCCCTTCAGAAGAACCA AATGAGGACCTGAAGCCGAGGTGAAAATCGAAGGTGGAACACAAATTCATCCAGTGTACA AATTCTGCAGCAGGTGTGAAGATCTTAACATCGTTCAAGTGAAGTGTGATAATGAGAAGGAA AGATTATCAAGCATTGAAAAGATTAAACAGCTAAGAGAACAAGTTAATGACCTCTTTAGCCGA AAATTTGGTAAAGCAATTGGCGTGGATTTCCCTGTGAAAGTCCCTACAGGAAGATCACATTC AACCCTGGCTGTGTGGTATTGATGGCATGCCCGGGGGTGGTATCAAGGCCCCGGCTAT CTGGAATCAGTTCATGAGGAGGATCTTGGAGGCAGCTGAGTTTATCAAATTCACAGTCATC AGGCCGCTTCCAGGGCTTGAGCTCAGTAATGTGGGAAAACGCAAGATAGACCAGGAGGCCGT GTGTTTCAAGAAAAGTGGGAGAGAGCGTATTTCTTGTGGAAGTACAGAATATTCACATGT CTCATATGCAACAAAGCATGTCTGTGTCCAAAGAATATAACCTAAGACGCCACTATCAAACC AATCACAGCAAGCATTATGACCAGTATATGGAAGAATGCGTGACGAGAAGCTTACAGAGCTG AAAAAAGGCTCAGGAAGTATCTTAGGCTTGTGACACCCAGTGTCCCGAGCAAAAACAA GTGTTTGC AAACCAAGTCCAACCCAGAAATCCCCGCTGCAGCCTGTAGAGGACCTAGCTGGG AACTTATGGGAGAAGTTACGTGAAAAAATCAGGTCTTTTGTGGCATATTCATCGCAATCGAT GAGATCACGGATATAAATAATACACCCAGTGTGGCATATTCATCCGTGGTGTGATGAGAAT TTGATGTGTCCGAAGAACTTCTGGACACGGTGCCCATGACGGGTACAAAATCTGGCAACGAG ATCTTTTCGCGTGTGTGGAAGAGCCTGAAAAAGTTCTGTATCGACTGGTGC AAATTAGTAAGC GTGGCCTCCACTGGCACCCAGCGATGGTGGATGCCAATAACGGGCTTGTACAAAAGTGAAG TCCAGGGTGGCGACGTTCTGCAAGGGTGC GGAAGTGAAGTCCATCTGTTGTATAATTATCCG GAATCACTCTGTGCTCAGAAGTTGAAGATGGACCAGTCTATGGACGTGGTGTGAGTCCGTG AACTGGATATGCTCCCGGGGACTGAACCACAGTGAAGTTCACAACCTTGCTCTATGAGCTGGAC AGCCAGTATGGTAGCCTCTGTACTACACGGAGATTAAGTGGCTCAGTCCGGGGCTCGTGCTA AAGAGATTTTTCGAATCCTTGAAGAAATCGACTCCTTCATGTATCCAGAGGGAACCCCTG CCTCAACTGAGCTCCATAGATTGGATCCGAGACCTGGCCTTCTTGGTTGACATGACGATGCAT CTGAACGCTTTGAACATCTCTCTCAAGGACACTCCCAATCGTCACGCAGATGTATGACCTG ATCCGGGCGTTCTAGCAAAAAGTGTGCTCTGGGAGACTCATTGACGAGGAATAATCTGGCC CACTTTCCACCCCTGAAATTGGCTTCCAGAAATGAAAGCGATGGCCTGAACTACATTCCTCAA ATCGCGGAAGTCAAGACCGAATTCAGAAAAGGCTGTCTGATTTCAAACCTACGAAAGCGAA CTGACTCTGTTTCAAGTCCCGCTTCTCCAGGAAGATCGACAGTGTGACGAGGAGCTCCAGAT GAGGTTATCGACCTGCAATGCAACACGGTCTTGAAGACGAAATACGACAGGATGGGAATACCA GAATTCACAAAGTACCTCTGGGGTAGCTACCCGAAATACAAGCACCATTGCGCAAGATTCTT TCCATGTTGCGGAGCCTACATTTGCGAACAGCTGTTCTCCATTATGAACTGAGCAAAACA AAATACTGCTCCAGTTAAAGGATTCAGTGGGATTCTGTACTCCACATCGCAACGTGATGG AGAGAAAAGTCTGGCAGGCGCTATGGTGGGAAAGGCTGGAGTCTTCTAGTCCCAAGGGATT GGGAGATGACAAAATGAATTTTCTTTTGA</p>
	<p>ORF Start: ATG at 163 ORF Stop: TGA at 3019</p>
	<p>SEQ ID NO: 298 952 aa MW at 107635.4kD</p>
NOV84c, CG59794-03 Protein Sequence	<p>MAQVAVSTLPVEESSSETRMVTFLVSALES MCKELAKSKAEVACIAYVETDVFVVGTERGC AFVNARTDFQKDFAKYCAEGLCEVKPPCPVNGMQVHSGETEILRKAVEDYFCFCYQKALGTT VMVPVPYEKMLRDQSAVVVQGLPEGVAFQHPENYDLATLKWILENKAGISFIINRRPFLGPES QLGGPGMVTDAERSIVSPSESCGPINVKTEPMEDSGISLKA EAVSVKKESEDPNYYQYNMQGN SHPSSTSNIEVEMELPMEQDSTPLVPSEEPNEDPEAEVKIEGGNTNSSSVTNSAAGVEDLNIV QVTVDNEKERLSSIEKIKQLREQVNDLFSRKFGKAIGVDFPVKVPYRKITFNPVCVVIDGMPP GVVFKAPGYLEISSMRRIEAAEFIKFTVIRPLPGLLESNVGRKIDQEGRVFQEKWERAYFF VEVQNIPTCLICKQSMVSKEYNLRRHYQTNHSHYDQYMERMRDEKLHELKGLRKYLLGLS DTECPQKQVFANPSPTQKSPVQPVEDLAGNLWEKLREKIRSFVAYSIAIDEITDINNTTQLA IFIRGVDFENFVSEELDTVPMTGKSGNEIFSRVEKSLKKFCIDWSKLVSVASTGTPAMVDA NNGLVTKLSRVATFCKGAELKSIICIIHPESLCAQKLKMDHVMVVKSVNWCISRLNHSE FTLLYELDSQYGSLLYYTEIKWLSRGLVLRKFESLEEIDSFMSRGRKPLPQLSLDWDRL AFLVDMTHMLNALNISLQGHSQIVTQMYDLIRAF LAKLCLWETHLTRNNLAHFPTLKLASRNE SDGLNYPKIAELKTEFQKRLSDFKLYESELTLFSSPFSTKIDSVHEELQMEVIDLQCNVTLK TKYDKVGIPEFYKYLWGSYPKYKHKHCAKILSMFGSTYICEQLFSIMKLSKTKYCSQLKDSQWD SVLHIAT</p>

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 84B.

Table 84B. Comparison of NOV84a against NOV84b and NOV84c.		
Protein Sequence	NOV84a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV84b	477..972 1..496	484/496 (97%) 484/496 (97%)
NOV84c	21..972 1..952	923/952 (96%) 923/952 (96%)

- 5 Further analysis of the NOV84a protein yielded the following properties shown in Table 84C.

Table 84C. Protein Sequence Properties NOV84a	
PSort analysis:	0.3600 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in lysosome (lumen); 0.1000 probability located in nucleus
SignalP analysis:	No Known Signal Sequence Predicted

- 10 A search of the NOV84a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 84D.

Table 84D. Geneseq Results for NOV84a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV84a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABG22337	Novel human diagnostic protein #22328 - Homo sapiens, 1205 aa. [WO200175067-A2, 11-OCT-2001]	19..972 315..1205	864/961 (89%) 871/961 (89%)	0.0
AAM79866	Human protein SEQ ID NO 3512 - Homo sapiens, 663 aa. [WO200157190-A2, 09-AUG-2001]	316..972 1..663	650/663 (98%) 653/663 (98%)	0.0
ABB11827	Human transcription factor 21-like protein homologue, SEQ ID NO:2197 - Homo sapiens, 663 aa. [WO200157188-A2, 09-AUG-2001]	316..972 1..663	650/663 (98%) 653/663 (98%)	0.0

AAM78882	Human protein SEQ ID NO 1544 - Homo sapiens, 582 aa. [WO200157190-A2, 09-AUG-2001]	396..972 1..582	575/582 (98%) 575/582 (98%)	0.0
AAB42755	Human ORFX ORF2519 polypeptide sequence SEQ ID NO:5038 - Homo sapiens, 533 aa. [WO200058473-A2, 05-OCT-2000]	440..972 1..533	531/533 (99%) 531/533 (99%)	0.0

In a BLAST search of public sequence databases, the NOV84a protein was found to have homology to the proteins shown in the BLASTP data in Table 84E.

Table 84E. Public BLASTP Results for NOV84a				
Protein Accession Number	Protein/Organism/Length	NOV84a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99NI3	Gtf2ird2 - Mus musculus (Mouse), 936 aa.	21..972 1..936	749/953 (78%) 836/953 (87%)	0.0
BAC04576	CDNA FLJ38253 fis, clone FCBBF3000768, moderately similar to Homo sapiens general transcription factor 2-I (GTF2I) mRNA - Homo sapiens (Human), 702 aa.	269..972 1..702	693/705 (98%) 697/705 (98%)	0.0
CAD38861	Hypothetical protein - Homo sapiens (Human), 496 aa.	477..972 1..496	495/496 (99%) 495/496 (99%)	0.0
CAD38788	Hypothetical protein - Homo sapiens (Human), 397 aa (fragment).	579..972 4..397	394/394 (100%) 394/394 (100%)	0.0
Q9H739	CDNA: FLJ21423 fis, clone COL04129 - Homo sapiens (Human), 364 aa.	609..972 1..364	360/364 (98%) 362/364 (98%)	0.0

5

PFam analysis predicts that the NOV84a protein contains the domains shown in the Table 84F.

Table 84F. Domain Analysis of NOV84a			
Pfam Domain	NOV84a Match Region	Identities/ Similarities for the Matched Region	Expect Value
GTF2I	127..203	38/77 (49%) 71/77 (92%)	1.1e-39
GTF2I	355..430	38/76 (50%) 72/76 (95%)	1.2e-41

Example 85.

The NOV85 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 85A.

Table 85A. NOV85 Sequence Analysis			
	SEQ ID NO: 299	530 bp	
NOV85a, CG59821-01 DNA Sequence	CAGGATGAACGCTGCTTTCCAAGATGGCGACGGAGGGAGGAGGGAAGGAGATGAACGAGATTA AGACCCAATTACACACCCGGGAAGGTCTGTACAAGCTGCTGCCGACTCGGAGTACAGCCGGC CCAACCGGGTGCCCTTCAACTCGCAGGGATCCAACCTGTCCGCGTCTCCTTCGTAAACCTCA ACGACCAGTCTGGCAACGGCGACCGCCTCTGCTTCAATGTGGGCCGGGAGCTGTACTTCTATA TCTACAAGGGGGTCCGCAAGGCTGCTGACTTGAGTAAACCAATAGATAAAAGGATATACAAAG GAACACAGCCTACTTGTTCATGACTTCAACCACCTAACAGCCACAGCAGAAAGTGTCTCTCTCC TAGTGGGCTTTTCCGAGGCCAAGTCCAGCTTATAGACCAATCAAAAAGAACTAGCAAAC TTTTAATGAGGAAGGCTCATTGTTCATCCCCAAGCCAGGCCAGTTCTCCAGGTGGAAGTGTAG GTAGCGACCTCACTGCTGCGCGCAC		
	ORF Start: ATG at 24		ORF Stop: TAG at 507
	SEQ ID NO: 300	161 aa	MW at 17692.7kD
NOV85a, CG59821-01 Protein Sequence	MATEGGKEMNEIKTQFTTREGLYKLLPHSEYSRPNRVFPNSQGSNPVRVSVFVNLNDQSGNGD RLCFNVGRELYFYIYKGVRAADLSKPIDKRIYKGTQPTCHDFNHLTATAESVSLLVGFSAGQ VQLIDPIKKETSKLFNEEGLSSPSQASSPGGTVV		
	SEQ ID NO: 301	519 bp	
NOV85b, CG59821-02 DNA Sequence	AGGATGAACGCTGCTTTCCAAGATGGCGACGGAGGGAGGAGGGAAGGAGATGAACGAGATTAA GACCCAATTACACACCCGGGAAGGTCTGTACAAGCTGCTGCCGACTCGGAGTACAGCCGGCC CAACCGGGTGCCCTTCAACTCGCAGGGATCCAACCTGTCCGCGTCTCCTTCGTAAACCTCAA CGACCAGTCTGGCAACGGCGACCGCCTCTGCTTCAATGTGGGCCGGGAGCTGTACTTCTATAT CTACAAGGGGGTCCGCAAGGCTGCTGACTTGAGTAAACCAATAGATAAAAGGATATACAAAGG AACACAGCCTACTTGTTCATGACTTCAACCACCTAACAGCCACAGCAGAAAGTGTCTCTCTCT AGTGGGCTTTTCCGAGGCCAAGTCCAGCTTATAGACCAATCAAAAAGAACTAGCAAAC TTTTAATGAGGAAGGCTCATTGTTCATCCCCAAGCCAGGCCAGTTCTCCAGGTGGAAGTGTAG GTAGCGACCTCACTG		
	ORF Start: ATG at 23		ORF Stop: TAG at 506
	SEQ ID NO: 302	161 aa	MW at 17692.7kD
NOV85b, CG59821-02 Protein Sequence	MATEGGKEMNEIKTQFTTREGLYKLLPHSEYSRPNRVFPNSQGSNPVRVSVFVNLNDQSGNGD RLCFNVGRELYFYIYKGVRAADLSKPIDKRIYKGTQPTCHDFNHLTATAESVSLLVGFSAGQ VQLIDPIKKETSKLFNEEGLSSPSQASSPGGTVV		

5

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 85B.

Table 85B. Comparison of NOV85a against NOV85b.		
Protein Sequence	NOV85a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV85b	1..161	147/161 (91%)
	1..161	147/161 (91%)

10

Further analysis of the NOV85a protein yielded the following properties shown in Table 85C.

Table 85C. Protein Sequence Properties NOV85a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV85a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several

5 homologous proteins shown in Table 85D.

Table 85D. Geneseq Results for NOV85a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV85a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB68534	Human GTP-binding associated protein #34 - Homo sapiens, 161 aa. [WO200105970-A2, 25-JAN-2001]	1..161 1..161	161/161 (100%) 161/161 (100%)	3e-90
ABG14369	Novel human diagnostic protein #14360 - Homo sapiens, 87 aa. [WO200175067-A2, 11-OCT-2001]	10..90 1..80	77/81 (95%) 78/81 (96%)	5e-39
AAM79336	Human protein SEQ ID NO 2982 - Homo sapiens, 687 aa. [WO200157190-A2, 09-AUG-2001]	12..144 21..221	89/201 (44%) 104/201 (51%)	2e-31
AAM78352	Human protein SEQ ID NO 1014 - Homo sapiens, 684 aa. [WO200157190-A2, 09-AUG-2001]	12..144 21..221	89/201 (44%) 104/201 (51%)	2e-31
ABG06239	Novel human diagnostic protein #6230 - Homo sapiens, 580 aa. [WO200175067-A2, 11-OCT-2001]	22..83 428..489	62/62 (100%) 62/62 (100%)	2e-31

In a BLAST search of public sequence databases, the NOV85a protein was found to have homology to the proteins shown in the BLASTP data in Table 85E.

Table 85E. Public BLASTP Results for NOV85a

Protein Accession Number	Protein/Organism/Length	NOV85a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9D721	2310040A13Rik protein - Mus musculus (Mouse), 161 aa.	1..161 1..161	156/161 (96%) 158/161 (97%)	4e-86
AAH30654	Hypothetical 22.2 kDa protein - Homo sapiens (Human), 195 aa.	1..146 1..146	145/146 (99%) 145/146 (99%)	1e-80
Q8TBZ3	Similar to putative - Homo sapiens (Human), 569 aa.	1..144 1..144	144/144 (100%) 144/144 (100%)	3e-80
AAH27497	Similar to RIKEN cDNA 2310040A13 gene - Mus musculus (Mouse), 145 aa.	1..144 1..144	142/144 (98%) 142/144 (98%)	6e-78
Q8R0J5	Similar to putative - Mus musculus (Mouse), 539 aa.	1..144 1..144	142/144 (98%) 142/144 (98%)	6e-78

PFam analysis predicts that the NOV85a protein contains the domains shown in the Table 85F.

Table 85F. Domain Analysis of NOV85a

Pfam Domain	NOV85a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 86.

The NOV86 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 86A.

Table 86A. NOV86 Sequence Analysis

	SEQ ID NO: 303	6064 bp	
NOV86a, CG59849-01 DNA Sequence	ATGACCACCAACGGAAATCATCGGCCCTCTGGTGCCATGCCGATGTTTCCGAGGTGAAGAA GAAATCATCTCAGTTTTAGATTACTCCCACTGCAGTCTTCAGCAGGTGCCAAAGGAGGTCTTT AACTTCGAACGAACATTAGAGGAGCTTTATCTAGATGCCAATCAAATTGAAGAACTACCCAAG CAATTGTTCAACTGTCAAGCTCTACGAAACTAAGTATTCTGATAACGACCTTTCAAATCTG CCAACCACTATTGCTAGTTAGTTAATCTTAAAGAACTCGACATCAGTAAAAATGGTGACAA GAATTTCCAGAAACATAAAGTGCTGTAAGTGTTTAAACAATTATTGAAGCCAGTGTCATCCC ATTTCTAAGCTACCTGATGGCTTCACACAGCTCCTAAACCTGACCCAGCTCTACCTGAATGAC GCCTTTCTTGAATTTCTTCCAGCCAATTTTGAAGGCTTGTCAAATTGCGGATCTTGGAGTTA AGAGAAAATCACTTGAAAACCTACCAAAGATGCACAACTGGCCCAAGTTGGAAAGACTTGAC CTAGGCAATAATGAATTCAGTGAGCTGCCTGAAGTTCTGGATCAAATACAAAATTTGAGGGAG TTATGGATGGATAATAATGCATTACAAGTGTTACCTGGGTCTATAGGGAAGTTAAAGATGTTG GTATACCTGGATATGTCAAAAACAGAATAGAAACAGTTGACATGGACATTTCTGGATGTGAA		

GCCCTTGAGGACCTCTTATTGTCATCCAATATGTTGCAACAATTGCCTGATTCTATAGGTGGA
CTTTTGAAAAAATAACAACCTCTAAAAGTAGATGACAATCAACTTACAATGCTACCCAATACA
ATTGGAAGTTTATCTTTATTAGAAGATTTGACTGTAGCTGTAATGAAGTGGAGTCACTACCT
TCTACTATTGGCTACCTTCATAGTCTTCGGACATTAGCAGTTGATGAGAATTTCTTCCAGAA
TTACCCAGAGAAATTGGAAGTTGTAAGAATGTAACAGTCATGTCTCTACGCTCCAACAATTA
GAATTTCTTCTGAAGAGATTGGACAGATGCAGAACTAAGAGTCTTAAATTTGAGTGACAAC
AGGTTGAAGAATTTACCATTTCTATTACCAAACTTAAAGAGCTTGACGCTTTGTGGCTTTCT
GACAATCAGTCCAAAGCCCTTATCCCTTTACAAACAGAAGCCCATCCAGAAACAAGCAAAGA
GTATTGACTAATACATGTTTCCCCAGCAGCCTCGTGGTGATGAAGATTTCCAGTCAGACAGT
GACAGCTTTAACCCTACACTGTGGGAAGAGCAGAGACAACAACGCATGACTGTTGCCTTTGAA
TTGAAGACAAAAAGAAGATGACGAAAATGCTGGGAAAGTTAAGCTCTCCTGCCAAGCCCC
TGGGAAAGGGGCCAGCGTGGGATTACTCTCCAACCTGCCAGACTGTCTGGCGATTGCTGCACA
CCATGGGCCAGGTGTGATCAGCAGATCCAAGATATGCCCGTCCCCCAGAAATGACCCAGCTG
GCATGGGGTTGTATAAGTGGCCTCCAGCAGGAAAGGAGCATGTGTACTCCATTGCCAGTTGCA
GCACAATCCACCCTCTTCCCTCTTAAGTGGCAGACAGGTTGAAATAAACCTAAAACGATAT
CCAACCTCCTTACCAGAGGATTTAAAGAATATGGTAAATCTGTTCAAATTTGGTGGGTAAAG
CCAAGCCATGGAGTGGTGTGAGAATTCAAATCCAATGCTAACACGGAGCAAATGTGAAA
GAAAAATATGAACACAAGTGGCCGGTAGCCCCAAAGGAGATTACAGTGGAGGATTCTTTGTT
CATCCAGCTAATGAAATGAGGATTGGGGAACCTTACCCTTCATTAGCTGAGACCCCTCTGTAC
CCACCCAACTTGTCTGTCTAGGGAAGGACAAAAAGAATCAACTGATGAGTCTGAAGTTGAC
AAAACCTACTGTCTGAATAACAGTGTTCCTCAGGCATCTACTCAGACTACTCGCCTTCCCAG
GCTTCCTCAGGATCCTCTAATACCGGGTTAAAGTGGGGTCTTGACAGACAACAGTAAGAT
GCAGTACATAATTCTTTGGGGTAACAGGATTGCACCATCTTTCCACAGCCTCTTGATTCA
AAGCCATTACTCAGCCAGCGGGAGGCTGTTCCTCCAGGCAATATACCACAGCGCTCTGACCGG
CTGCCCATGAGTGATACTTTCACTGACAACCTGGACTGATGGCTCGCATTATGACAACACAGGG
TTTGTGTGCTGAGGAAACACAGCCGAGAATGCCAACAGTAATCCTCTCTTAAGTTCGAAATCT
AGAAGCACATCTTCGCATGGACGCGAGGCCCTTTGATCAGGCAAGACAGGATTGTGGTGTCCC
CTGGAACCTCGAGCAGTCTACACACAGACACACACCAGAAACAGAAGTGCCTCCTTCCAATCCT
TGGCAGAAATTGGACCAGAACCCCTAGTCCGTTTGAAGACAGGACCGCTTTTCTCTCAAATTA
GAGACAACCCCCACTACCAGCCCATTTGCCTGAAAGGAAAGACATATAAAGGAATCTACTGAA
ATACCTAGTCCCTTTTCTCAGGCGTACCATGGGAGTATCATGATTCCAATCCCAACAGGAGT
CTTAGTAATGTCTTTTCTCAAATCCATTGCGCGCCGAATCTTCTAAAGGTGTTATTTCATTT
AGCAAAAGCACAGAGAGGCTTTCCCCCTAATGAAAGATATCAAGTCTAATAAATTCAAAAAG
TCACAGAGTATCGATGAGATTGACATTGGTACATATAAGGTGTATAACATACCATTAGAAAAAC
TATGCTTCTGGGAGTGATCACTTAGGAAGCCACGAACGACCGGATAAGATGCTGGGACAGAG
CATGGTATGTCCAGTATGTCTCGAAGCCAGTCAGTCCCAATGCTGGATGATGAGATGCTCACC
TACGGAAGTAGTAAGGGGCCACAACAACAAAAAGCTTCTATGACAAAAAAGTCTATCAGTTT
GACCAAGCTTCAATCCTCAAGGATCAGTGGAAAGTGAAGCCGAAAAAGAGGATACCACCCCT
TTTCAACACAATCCCGAGTACGTGCAACAGGCCAGCAAAAACATCGCAAGGATTGATTAGT
CCTAGAGCTTACAGAGGATACCCACCGATGGAGCAAAATGTTTCTTCTCAGCCATCTGTG
AATGAGGATGCTGTGGTGAATGCCAGTTCGCAAGCCAAGGGGCCAGGGCGGGCTTCTGTAGA
AGGGCCGACTCCCTGGTGAGCGCCACAGAAATGGCCATGTTTAGAAGGGTCAATGAGCTCAT
GAGCTGCCCCCACTGATAGGTACGGCAGACCCCATATAGGGGAGGGTGGATCGCCAAAGC
AGCGTTACAGTGACTGAGTCCAGTTCCTGAAAAGGAATGGCAGGTATGAAGATGAACACCCT
TCATATCAAGAAGTGAAAGCTCAGGCGGGAAGTTTCCGGTTAAAAACCTTACCAAAAGGAGG
CCATTGTCTGCGAGAAGCTACAGTACAGAGAGTTACGGTGCCTCCCAAAACAGGCCAGTTTCA
GCTAGGCCTACTATGGCAGCTCTTTTGGAAAAATACCATCTGACTATAACTTGGGTAACTAT
GGTGACAAGCCATCAGATAACAGTGATTTAAAGACGAGGCCCTACTCCTGTGAAGGGAGAGGAG
AGCTGTGGTAAATGCCTGCAGACTGGAGACAACAGCTGCTTAGACATATAGAAGCTAGACGG
TTAGACAGGACCCCGTCCAGCAAAGCAACATTTTAGACAATGGACAAGAAGATGTATCTCCT
AGTGGCCAATGGAATCCTTATCCACTTGGGAGGCGGGATGTACCTCCGGACACCATTACTAAG
AAGGCAGGCAGCCACATCCAGACGTTGATGGGGTCCCAAGCCTTTCAGCATCGCAGCCGGGAG
CAGCAGCCGTATGAAGGAATATATAACAAGTGACCATCCAGCAATTCAGTCACCATGTGCT
ATTAGATCCCTCTTACAGGCCACCCGGGACCTCAGCCTGGACGGTGTCTTAATTCAAACCT
AAAGGGCAAAGGAGTATGGATGGATATCCAGAGCAGTTTGTGTGAGAATAGAAAAGAAATCCT
GGCCTTGGATTAGTATCAGTGGTGAATTAGTGGACAAGGAAATCCATTCAAACCTTCTGAC
AAGGGTATCTTTGTTACTAGGGTTCAGCCTGATGGGCCAGCATCAAACCTACTGCAGCCTGGT
GATAAGATCCTTCAGGCAATGGACACAGTTTGTACATATGGAACATGAAAAGCTGTATTATA
CTACTGAAGAGTTTCCAGAACACAGTAGACCTAGTTATTCAACGTGAGCTTACTGTCTAAATA
TTTTTTATAAATAGTGAAGATACGTCTAGCCAGACCTAATGTTCAAAAATAAATTTATACATA
GAAACAATTTTGCCAATTGCTGGACCAATGGCAACATTAGTGCCAATGTATAATACTATA
TGTTAGCACTGACCATCCTTAAAAATGTTAACTCTATAAATATGATGTTTCATGTGGTTATGT
ATTAGTTTAAATTTGTGACGCTCTGGCTGTGCATTGGTGCAGTTTGTCTTCTTTT
GTTTTTAATCAAATAAGTTTCTTCTCAAATGGATTTTATATAATTTCGGAGCACGGAAGCAC
ACACAAGCTCTTTATGAATTTCTGCTCTCCATCAGAAACACTGCCTCAAAGTTGTATATGCCTT
TATATAGAAAATACAAATATAAGAATTGTAATCCCATAAAATATTTCTAGCACAAAGGTATA

	<p>TGTTGGCATATATACAAAAAGAATATAGAGAAAAACAATATTTTCATAAACTAAACATCTCAG ATAGAGAAAAATATATCTTAAATAAGACTTTACTATATTGAATCTTTTCAATAAAATTA CATGATAATGCCTTATGAAAGTAAGCTGTACATATGGTATAAAGTGTTTATATTTGGTTCCATA TTCATTTGCTAAATCTCATGACACAGAGTGAAATATTTTCATAAAATAGCCATTTATCTCTGG GACCCAAATAAAATAGGATGAACATAATTGTTCAATGCCTTTAGCTAATTACAATACATGCA GAGTTTAGAAACAGACTAAAGGTCAATTGTAGTTAAGTCTTTTACCACAAATTTAAGCAGTG GATGATGGGTGGCAGGAAAGGTATGCTTTATTTCTTCAAGTTCATGTTGATTATAAACTGT AGCCCCGTGATTCTTTACTTGTAAATGTGAATTTATTTGTGTGTGCTTAATCTAATTTG CTGCTTTTAAATTATTTAAACGAATTTTGGAAATTGATAAAATTTATCATTACGAAAGACT GCTGTTAGAAAGTTATGGTAGGTGATTTTAAATCCTTGGTATTTAAATATGAAACTTCAATA TAATTTCTCAGAGCTGTGGTCTACCTGTATCATTAAATTTCAATGGCTGTTTTCTGGGCAGAA ATAGATAAAATACTTTTTCCAAAAACAGTTTCAAGGTATGTAAAACTCGTAATGCTTTTTT ACTGAAGAGAAAGACAAGCATGGTTAATGTAGAATTATTTACTTTTCCATTGAAACTATTTTC CTGCATAAATGATCAAAATTTATTTTATAATCCTTTAAATACTTATCTTTTCATATTAGTCAT TAATTTAATTACAATATTAATTTGAATTTCCAGGATAATTTCCCGGAGTTGGTTCATGCATT ATCTTTTATAATTTACATAGTTCTTTTGTATATAATGAATTTACTTTACATGCTAGTGT CAAGTATGTATGAGGATTTTCACAATAGTATCACTGAATGATGTACCAGAGCTCTGAGAAT AATATTTGTAAGTTAACTGTTTTATGGGGACATTGAAAAATATGTATTTTGTAGGGTCTATT AAAATGAGTGTCACTT</p>
	<p>ORF Start: ATG at 1</p> <p>ORF Stop: TAA at 4468 ,</p>
	<p>SEQ ID NO: 304</p> <p>1489 aa</p> <p>MW at 167241.9kD</p>
NOV86a, CG59849-01 Protein Sequence	<p>MTTKRKII GRLVPCRCFRGEEII ISVLDYSHCSLQQVPKEVNFERTLEELYLDANQIEELPK QLFNCQALRKLSIPDNDLSNLPPTIASLVNLKELDISKNGVQEPENIKCKCLTIIEASVNP ISKLPDGTQLNLTLQYLNDALFLPANFGRVLKRLILELRENHLKTLPMKMKLAQLERLD LGNNEFSSELPEVLDQIQNRELWMDNNAQVLPGSIGKLMVLVLDMSKNRIETVDMDISGCE ALEDLLSSNMLQQLPDSIGLLKLTLLKVDNQLTMLPNTIGSLSLLEEFDCSCNELESPL STIGYLHSLRTLAVDENFLPELPREIGSCKNVTVMSLRSNKLEFLPEEIGQMOKLRLVNLSDN RLKNLPFSFTKLKELALWLSDNQSKALIPLQTEAHPETKQRVLTNYMFPQQRGDEDFQSDS DSFNPTLWEEQQRMTVAFEFEDKKEDDENAGKVKLSQCAPWERGQGITLQPARLSGDCCT PWARCDQIQDMPVPQNDPQLAWGCISGLQQRSMCTPLPVAQSTTLPSLSGRQVEINLKRY PTYPEDLKNMVKSVQNLVGKPSHGVRVENSNPNTANTEQTVKEKYEHKWPVAPKEITVEDSFV HPANEMRIGELHPSLAETPLYPKLVLLGKDKKESTDESEVDKTHCLNNSVSSGTYSDYSPSQ ASSGSSNTRVKVGLQTAKDAVHNSLWGNRIAPSFPPQLDSKPLLSQREAVPPGNIPQRPDR LPMSDTFTDNWTDGSHYDNTGFVAEETTAENANSNPLSSKSRSTSSHGRRLPIRQDRIVGVP LELEQSTHRTPETEVPPSNPWQNWTRTPSPFEDRTAFPSKLETTPTTSPLPERKEHIKESTE IPSPFSPGVWPWEYHDSNPNRSLSNVFSQIHCPRRESSKGVISISKSTERLSPLMKDKIKSNKFKK SQSIDEIDIGTYKVYNIPLNYASGSDHLGSHERPDKMLGPEHGMSSMSRSQSVPMDDDEMLT YGSSKGPQQQKASMTKKVYQFDQSFNPQGSVEVKAERKIPPFQHNPEYVQASKNIAKDLIS PRAYRGYPPEQMFSFSQPSVNEDAVVNAQFASQGARAGFLRRADSLVSATEMAMFRRVNEPH ELPPTDRYGRPPYRGGLDRQSSVTVTESQFLKRNTRYEDHPSYQEVKAQGSFPVKNLQRR PLSARSYSTESYGASQTRPVSARPTMAALLEKIPSDYNLGNYGDKPSDNSDLKTRPTPVKGEE SCGKMPADWRQQLLRHIEARRLDRTPSQQSNILDNGQEDVSPSGQWNPYPLGRRDVPDPTITK KAGSHIQTLMSQSLSQHSRREQPYEGNINKVTIQFQSPPLPIQIPSSQATRGPQPGRCCLIQT KGQSRMDGYPEQFCVRIEKNPGLGFSISGGISGQGNPFKPSDKGIFVTRVQPDGPASNLLQPG DKILQANGHSFVHMEHEKAVLLLSKFQNTVDLVIQRELTV</p>
	<p>SEQ ID NO: 305</p> <p>1260 bp</p>
NOV86b, CG59849-02 DNA Sequence	<p>ATGACCACCAAACGGAATCATCGGCCGTCTGGTGCCATGCCGATGTTTCCGAGGTGAAGAA GAAATCATCGCAGTTTATGATTACTCCACTGCAGTCTTCAGCAGGTGCCAAAGGAGGTCTTT AACTTCGAACGAACATTAGAGGAGCTTTATCTAGATGCCAATCAAATTGAAGAACTACCCAAG CAATTGTTCAACTGTCAAGCTCTACGAAAACTAAGTATTCCTGATAACGACCTTTCAAATCTG CCAACCACTATTGCTAGTTAGTTAATCTTAAAGAACTCGACATCAGTAAAAATGGTGTACAA GAATTTCCAGAAAAACATAAAGTCTGTAAGTGTTAACAAATTATTGAAGCCAGTGCTACATCC ATTTCTAAACTACCTGATGGCTTCACACAGCTCCTAAACCTGACCCAGCTCTACCTGAATGAC GCCTTCTTGAATTTCTTCCAGCCAATTTGGAAGACTTGTCAAATTGCGGATCTTGGAGTTA AGAGAAAACTACTTGAAACTCTACCAAAGTCAATGCACAACTGGCCAGTGTGGAAGAGT GACCTAGGCAATAATGAATTCGGTGAGCTGCCTGAAGTTCTGGATCAAATACAAATTTAGG GAGTTATGGATGGATAATAATGCATTACAAGTGTACCTGGGGCAGGCAGCCACATCCAGACG TTGATGGGGTCCCAAAGCCTTACGATCGCAGCCGGGAGCAGCAGCCGTATGAAGGAAATATA AACAAAGTGACCATCCAGCAATTCAGTCACCATTTGCCTATTAGATCCCTCTTACAGGGCC ACCCGGGGACCTCAGCCTGGACGGTGTAAATCAAATAAAGGGCAAAGGAGTATGGATGGA TATCCAGAGCAGTTTGTGTGAGATAAGAAAAGAAATCCTGGCCTTGATTAGTATCAGTGGT GGAATTAGTGGACAAGGAAATCCATTCAAACCTTCTGACAAGGGTATCTTGTACTAGGGTT CAGCCTGATGGCCAGCATCAAACCTACTGCAGCCTGGTGATAAGATCCTTCAGGCAAATGGA CACAGTTTGTACATATGGAACATGAAAAGCTGTATTACTACCGAAGAGTTTCCAGAACACA</p>

	GTAGACCTAGTTATTCAACGTGAGCTTACTGTCTAAATATTTTATAAAATAGTGAAGATACG TCTAGCCAGACCTAATGTTCAAAAATAAATTTATACATAGAAACAAATTTGCCAATTGCTGG		
	ORF Start: ATG at 1		ORF Stop: TAA at 1168
	SEQ ID NO: 306	389 aa	MW at 43792.9kD
NOV86b, CG59849-02 Protein Sequence	MTTKRKIIGRLVPCRCFRGEEIIAVLDYSHCSLQQVPKEVFNFFERTLEELYLDANQIEELPK QLFNCQALRKLSIPDNDLSNLPPTIASLVNLKELDISKNGVQEFFPENIKCKCLTIEASVNP ISKLPDGFQLLNLTQLYLNDALFLEFLPANFGRLVKLRILELRENHLKTLPKSMHKLAQLERL DLGNNEFGELPEVLDQIQNLRELWMDNNALQVLPAGAGSHIQTLMGSSQLQHREREQQPYEGNI NKVTIQFQSPPLPIQIPSSQATRGPPGRCLIQNKQORSMDGYEQFCVRIEKNPGLGFSISG GISGQGNPFKPSDKGIFVTRVQPDGPASNLLQPGDKILQANGHSFVHMEHEKAVLLPKSFQNT VDLVIQRELTV		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 86B.

Table 86B. Comparison of NOV86a against NOV86b.		
Protein Sequence	NOV86a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV86b	1..232	210/233 (90%)
	1..233	215/233 (92%)

5

Further analysis of the NOV86a protein yielded the following properties shown in Table 86C.

Table 86C. Protein Sequence Properties NOV86a	
PSort analysis:	0.5192 probability located in mitochondrial matrix space; 0.3000 probability located in microbody (peroxisome); 0.2487 probability located in mitochondrial inner membrane; 0.2487 probability located in mitochondrial intermembrane space
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV86a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 86D.

Table 86D. Geneseq Results for NOV86a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV86a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM52529	Human Erbin mutein #5 - Homo sapiens, 1371 aa. [FR2807437-A1, 12-OCT-2001]	1..1488 1..1370	566/1557 (36%) 790/1557 (50%)	0.0
AAM52528	Human Erbin mutein #4 - Homo sapiens, 1371 aa. [FR2807437-A1, 12-OCT-2001]	1..1488 1..1370	565/1557 (36%) 788/1557 (50%)	0.0
AAM52530	Human Erbin mutein #6 - Homo sapiens, 1371 aa. [FR2807437-A1, 12-OCT-2001]	1..1488 1..1370	565/1557 (36%) 787/1557 (50%)	0.0
AAM52527	Human Erbin mutein #3 - Homo sapiens, 1371 aa. [FR2807437-A1, 12-OCT-2001]	1..1488 1..1370	566/1557 (36%) 788/1557 (50%)	0.0
AAM52526	Human Erbin mutein #2 - Homo sapiens, 1419 aa. [FR2807437-A1, 12-OCT-2001]	1..1488 1..1418	568/1579 (35%) 793/1579 (49%)	0.0

In a BLAST search of public sequence databases, the NOV86a protein was found to have homology to the proteins shown in the BLASTP data in Table 86E.

Table 86E. Public BLASTP Results for NOV86a				
Protein Accession Number	Protein/Organism/Length	NOV86a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q96NW7	Densin-180 - Homo sapiens (Human), 1537 aa.	1..1489 1..1537	1486/1538 (96%) 1487/1538 (96%)	0.0
P70587	Densin-180 - Rattus norvegicus (Rat), 1495 aa.	1..1489 6..1495	1421/1491 (95%) 1454/1491 (97%)	0.0
Q9P2I2	KIAA1365 protein - Homo sapiens (Human), 831 aa (fragment).	659..1489 1..831	829/831 (99%) 830/831 (99%)	0.0
Q96RT1	Densin-180-like protein - Homo sapiens (Human), 1412 aa.	1..1488 1..1411	573/1562 (36%) 804/1562 (50%)	0.0
Q9NR18	ErbB2-interacting protein ERBIN - Homo sapiens (Human), 1371 aa.	1..1488 1..1370	567/1557 (36%) 789/1557 (50%)	0.0

5

PFam analysis predicts that the NOV86a protein contains the domains shown in the Table 86F.

Table 86F. Domain Analysis of NOV86a			
Pfam Domain	NOV86a Match Region	Identities/ Similarities for the Matched Region	Expect Value
LRR	47..69	9/25 (36%) 19/25 (76%)	0.13
LRR	93..115	8/25 (32%) 19/25 (76%)	0.83
LRR	184..206	10/25 (40%) 19/25 (76%)	0.048
LRR	207..229	12/25 (48%) 19/25 (76%)	0.041
LRR	253..275	12/25 (48%) 18/25 (72%)	0.71
LRR	277..299	8/25 (32%) 21/25 (84%)	0.13
LRR	369..391	11/25 (44%) 20/25 (80%)	0.00084
PDZ	1400..1486	34/93 (37%) 74/93 (80%)	8.5e-19

Example 87.

The NOV87 clone was analyzed, and the nucleotide and encoded polypeptide
 5 sequences are shown in Table 87A.

Table 87A. NOV87 Sequence Analysis			
	SEQ ID NO: 307	2062 bp	
NOV87a, CG59920-01 DNA Sequence	GAGGGGACGTCGTCGTAGAGGGCCGGAGCGGGCGGGCGGCGACGGACCCGGCTCCCGCGCAGG ACGGAGCCGCTGGCTCAGGTCGGGCCCTCCCCAACACCACCCGGGCCCTCCGCCCTTCCTGGG CCTCTCGGTGGAGCAGGGACCCGAACCGGTGCCCATCCAGTCCGGTGCCATCTGAAGCCCCCT TCCCAGAAAATGAGCCACAGAGCAAGCTGACCCAGCGACACAGCCCCCAGCCCTACTATAT TTCCGTTCTATCAAAAAATGGATGACTCGGAGACAGGTTTCAATCTGAAAGTCGTCTGGTC AGTTTCAAGCAGTGTCTCGATGAGAAGGAAGAGGTCTTGCTGGACCCCTACATTGCCAGCTGG AAGGGCTTGGTCAGGTTTCTGAACAGCCTGGGCACCATCTTCTCATTATCTCCAAGGACGTG GTCTCCAAGCTGCGGATCATGGAGCGCCTCAGGGGCGGCCCGCAGAGCGAGCACTACCGCAGC CTGCAGGCCATGGTGGCCACGAGCTGAGCAACCGGCTGGTGGACCTGGAGGGCCGCTCCAC CACCCGGAGTCTGGCTGCCGGACGGTGTGCGCCTGCACCGCGCCCTGCACTGGCTGCAGCTG TTCCTGGAGGGCCTGCGTACCAGCCCCGAGGACGCACGCACCTCCGCACTCTGCGCCGACTCC TACAACGCCTCGCTGGCCGCTACCACCCCTGGGTCTGCGCCGTGCCGTACCGTGGCCTTC TGCACGCTGCCACACGCGAGGTCTTCTGGAGGCCATGAACGTGGGGCCCCCGAGCAGGCC GTGCAGATGTAGGCGAGGCCCTCCCTTCATCCAGCGTGTCTACAACGTCTCCAGAAGCTC TACGCCGAGCACTCCCTGTCTGGACCTGCCCTAGAGGCGGGAAGCCAGGCGCGACCGCTTTC CTGCTGCAGATCTGGGCTGCGGTGGCCAGGGCCGTGAGTCCCGTGGCAGAGCCTTCTGGGCGC TGCGGGAACAGGAGATCCTCTGTGCCCCCTGTGAGCTGAGCTGGTTAGGAACACAGACTGTG ACAGAGAAGGTGGCGACCCAGCCAGAGAGGCCACCCCTCTCGGTCCGGAACAAGACGCCTCA GCCACGGCTCCCCCTCGGCCTATTACACGCGTGCGCAGCCAGGCCTCGCCAGGGTGCGGTGCA		

	GAGCAGAGCAGGCAGGGGTGGGGGCCGGGCCCAAGAGCCCGAAAGGTCGCCACCCCTAGC CTGTGGGGTGCATCTGCGAACCAGGGTGAAGTACAGGTCCCGGGGTGTGGAGGCTCCATCCT TTCTCCTTTCTGCCAGCCGATGTGTCTCATCTCAGGCCCGTGCCTGGGACCCCGTGTCTGCC CAGGTGGGCAGCCTTGAGCCAGGGGACTCAGTGCCCTCCATGCCCTGGCTGGCAGAAACCT CAACAGCAGTCTGGGCACTGTGGGGCTCTCCCGCCTCTCCTGCCTTGTTGCCCTCAGCGT GCCAGGCAGACTGGGGCAGGACAGCCGGAAGCTGAGACCAAGGCTCCTCACAGAAGGGCCCA GGAAGTCCCGCCCTTGGGACAGCCTCCTCCGTAGCCCTGCACGGCACCAGTTCCTCCGAGGG ACGCAGCAGGCCCGCTCCCGCAGCGCCGTGGGTCTGCACAGCCAGCCAGCCCAAGGCCCC CAGGAGCTGGGACTCTGCTACACCCAGTGAAATGCTGTGTCCCTTCTCCCGTGCCTTGA TGCCCCCTCCCAACAGTGCTCAGGAGACCCGTGGGGCAGGAACAGGAGGCTTGGACCTGT GGCCAGCCAAAGGCTACCAGACAGCCACAACAGCCAGCCACCATCCAGTGCCTGGGGCCT GGCCACTGGCTCTTACAGTGGACCCAGCACCTCGGGGTGGCAGAGGGACGGCCCCACGGC CCAGCAGACATGCGAGCTTCAGAGTGAATCTATGTGATGTCTTCAACGTTAATAAATCAC ACAGCCTCCAGGAGGGAGACGCTGGGGTGCAAAAAAAGCAAAA		
	ORF Start: ATG at 271 ORF Stop: TAG at 913		
	SEQ ID NO: 308	214 aa	MW at 24265.6kD
NOV87a, CG59920-01 Protein Sequence	MDDSETGFNLKVVLVSFKQCLDEKEEVLLDPYIASWKGLVRFNLGLTIFSFISKDVSKLRI MERLRGGPQSEHYRSLQAMVAHELNRNLVDLEGRSHHPESGCRVTLRLHRLHQLFLEGLR TSPEDARTSALCADSYNASLAAYHPVVRRAVTVAFTLPTRVFLFLEAMNVGPPEQAVQMLGE ALPFIQRVYNVSQKLYAEHSLDLLP		

NOV87b, CG59920-02 DNA Sequence	SEQ ID NO: 309 723 bp		
	ACAGCCCCCAGCCCTACTGTATTTCCGTTCCTATCAAAAAATGGATGACTCGGAGACAGGTT TCAATCTGAAAGTCGTCCTGGTCAGTTTCAAGCAGTGCTCGATGAGAGGAAGAGGCTTGC TGGACCCCTACATTGCCAGCTGGAAGGGCCTGGTCAGGTTTCTGAACAGCCTGGGCACCATCT TCTCATTATCTTCAAGGACCTGGTCTCAAGCTGCGGATCATGGAGCGGCTCAGGGCGGCC CGCAGAGCGAGCACTACCGCAGCCTGCAGGCCATGGTGGCCACGAGCTGAGCAACCGGCTGG TGGACCTGGAGCGCGCTCCCAACCCCGAGTCTGGCTGCCGACGGTGTGCGCCTGCACC GCGCCCTGCAGTGGCTGCAGTGTCTCTGGAGGGCCTGCGTACCAGCCCCGAGGACGCACGCA CCTCCGCGCTCTGCGCCGACTCCTACAACGCTCGTGGCGCCTACACCCCTGGGTCGTGC GCCGTGCCGTACCGTGGCCTTCTGCACGCTGCCCACACGCGAGGTCTTCTGGAGGCCATGA ACGTGGGGCCCCCGAGCAGGCCGTGCAGATGCTAGGCGAGGCCCTCCCTTATCCAGCGTG TCTACAACGTCTCCAGAAGCTCTACGCCGAGCACTCCCTGTGGACCTGCCCTAGAGCGGG AAGCCAGGGCCGCACCGCTTCTCTGCTGC		
	ORF Start: ATG at 42 ORF Stop: TAG at 684		
	SEQ ID NO: 310	214 aa	MW at 24364.8kD
NOV87b, CG59920-02 Protein Sequence	MDDSETGFNLKVVLVSFKQCLDEKEEVLLDPYIASWKGLVRFNLGLTIFSFISKDVSKLRI MERLRGGPQSEHYRSLQAMVAHELNRNLVDLERRSHHPESGCRVTLRLHRLHQLFLEGLR TSPEDARTSALCADSYNASLAAYHPVVRRAVTVAFTLPTRVFLFLEAMNVGPPEQAVQMLGE ALPFIQRVYNVSQKLYAEHSLDLLP		
NOV87c, 277583351 DNA Sequence	SEQ ID NO: 311 664 bp		
	CACCGGATCCACCATGGATGACTCGGAGACAGGTTTCAATCTGAAAGTCGTCCTGGTCAGTTTC AAGCAGTGTCTCGATGAGAAGGAAGAGGTTCTGTGGACCCCTACATTGCCAGCTGGAAGGGCC TGGTCAGGTTTCTGAACAGCCTGGGCACCATCTTCTCATTATCTTCAAGGACGTGGTCTCCAA GCTGCGGATCATGGAGCGCCTCAGGGCGGCCCGCAGAGCGAGCACTACCGCAGCCTGCAGGCC ATGGTGGCCACGAGCTGAGCAACCGCTGGTGGACCTGGAGCGCGCTCCACACCCCGAGT CTGGCTGCCGACGGTGTGCGCTGCACCGCGCCCTGCAGTGGCTGCAGTGTCTTCTGGAGGG CCTGCGTACCAGCCCCGAGGACGCACGACCTCCGCGCTCTGCGCCGACTCCTACAACGCTCG CTGGCCGCTACACCCCTGGGTGCTGCGCGCTGCGTACCGTGGCCTTCTGCACGCTGCCCA CAGCGAGGTCTTCTGGAGGCCATGAACGTGGGGCCCCCGAGCAGGCCGTGCAGATGCTAGG CGAGGCCCTCCCTTATCCAGCGTGTCTACAACGTCTCCAGAAGCTCTACGCCGAGCACTCC CTGCTGGACCTGCCCTCGAGGGC		
	ORF Start: at 2 ORF Stop: end of sequence		
	SEQ ID NO: 312	221 aa	MW at 25010.4kD
NOV87c, 277583351 Protein Sequence	TGSTMDDSETGFNLKVVLVSFKQCLDEKEEVLLDPYIASWKGLVRFNLGLTIFSFISKDVSKLRI KLRIEMERLRGGPQSEHYRSLQAMVAHELNRNLVDLERRSHHPESGCRVTLRLHRLHQLFLEGLR EGLRTSPEDARTSALCADSYNASLAAYHPVVRRAVTVAFTLPTRVFLFLEAMNVGPPEQAVQ MLGEALPFIQRVYNVSQKLYAEHSLDLLPLEG		

	SEQ ID NO: 313	2062 bp	
NOV87d, CG59920-01 DNA Sequence	GAGGGGACGTCGTCGTAGAGGGCCGAGCGGGCGGGCGGACCGGCTCCCGCGCAGG ACGGAGCCGTGGCTCAGGTCGGGCCCTCCCCAACACACCCCGGGCTCCGCCCCCTCTGGG CCTCTCGGTGGAGCAGGGACCCGAACCGGTGCCATCCAGTCCGGTGCCATCTGAAGCCCCCT TCCCAGAAAATGAGCCACAGAGCAAGCTGACCCACAGCAGACAGCCCCCAGCCCCACTATAT TTCCGTTCTATCAAAAAATGGATGACTCGGAGACAGGTTCAATCTGAAAGTCGTCTGGTC AGTTTCAAGCAGTGTCTCGATGAGAAGGAAGAGGTCTTGCTGGACCCCTACATTGCCAGCTGG AAGGGCCTGGTCAGGTTTCTGAACAGCCTGGGCACCATCTTCTATTCTCAAGGACGTG GTCTCCAAGCTGCGGATCATGGAGCGCCTCAGGGGCGGCCGAGAGCGAGCACTACCGCAGC CTGCAGGCCATGGTGGCCACAGAGCTGAGCAACCGGCTGGTGGACCTGGAGGGCCGCTCCCAC CACCCGAGTCTGGCTGCGGACGGTGCTGCGCCTGCACCGCGCCCTGACTGGCTGCAGCTG TTCCTGGAGGGCTGCGTACCAGCCCCGAGGACGACGACCTCCGCACTTGCGCCGACTCC TACAACGCTCGCTGGCCGCTACCACCCCTGGGTCGTGCGCCGTGCCGTACCGTGGCCTTC TGCACGCTGCCACACGCGAGGTCTTCTGGAGGCCATGAACGTGGGGCCCCCGAGCAGGCC GTGCAGATGCTAGGCGAGGCCCTCCCTTCATCCAGCGTGTCTACAACGTCTCCAGAAGCTC TACGCCGAGCACTCCCTGCTGGACCTGCCCTAGAGGCGGGAAGCCAGGGCCGACCGGCTTTC CTGCTGCAGATCTGGGCTGCGGTGGCCAGGGCCGTGAGTCCCGTGGCAGAGCCTTCTGGGCGC TGCGGGAACAGGAGATCTCTGTGCCCCCTGTGAGCTGAGCTGGTTAGGAACACAGACTGTG ACAGAGAAGGTGGCGACCCAGCCAGAAGAGGGCCACCCTCTCGGTCCGGAACAAGACGCTCA GCCACGGCTCCCGCTCGGCTATTACACGCTGCGCAGCCAGGCTCGCCAGGTGCGGTGCA GAGCAGAGCAGCGAGGGTGGGGCCGGGCCCGCAAGAGCCGAAAGGTCCGCCACCCCTAGC CTGTGGGTGCATCTGCGAACCAGGGTGAAGTCAAGTCCCGGGGTGTGGAGGCTCCATCCT TTCTCCTTTCTGCCAGCGATGTGTCTCATCTCAGGCCCGTGCCTGGGACCCCGTGTCTGCC CAGGTGGGCAGCCTTGAGCCAGGGGACTCAGTGCCCTCCATGCCCTGGCTGGCAGAAACCTC CAACAGCAGTCTGGGCACTGTGGGCTCTCCCGCCTCTCTGCTGTGTTTGGCCCTCAGCTG GCCAGGCAGACTGGGGGAGGACAGCCGGAAGCTGAGACCAAGGCTCCTCACAGAAGGGCCCA GGAAGTCCCGCCCTTGGGACAGCCTCCTCCGTAGCCCTGCACGGCACCAGTTCCCCGAGGG ACGCAGCAGGCCGCTCCCGCAGCGGCCGTGGGTCTGCACAGCCAGCCAGCCCAAGGCCCC CAGGAGCTGGGACTCTGCTACACCCAGTGAAATGCTGTGCTCCCTTCCCCGCTGCCCTTGA TGCCCCCTCCCAACAGTGTCTCAGGAGACCCGTGGGGCAGGAACAGGAGGTCTGGACCTGT GGCCAGCCAAAGGCTACCAGACAGCCACAACCAGCCAGCCACCATCCAGTGCCTGGGGCCT GGCCACTGGCTCTTACAGTGGACCCAGCACCTCGGGGTGGCAGAGGGACGGCCCCACGGC CCAGCAGACATCGAGCTTCCAGAGTGCAATCTATGTGATGTCTTCAACGTTAATAATCAC ACAGCTCCAGGAGGAGACGTGGGGTGCAAAAAAAGCAAAA		
	ORF Start: ATG at 271		ORF Stop: TAG at 913
	SEQ ID NO: 314	214 aa	MW at 24265.6kD
NOV87d, CG59920-01 Protein Sequence	MDDSETGFNLKVVLVSFKQLDEKEEVLDPYIASWKGLVRFNLSTGTFISFKDVSFKLRI MERLRGGPQSEHYRSLQAMVAHELNSRLVDLEGRSHHPESGCRTVLRLHRLHWLQFLGLELR TSPEDARTSALCADSYNASLAAYHPWVVRRAVTVAFCTLPTRVFLFAMNVGPPEQAVQMLGE ALPFIQRVYVNSQKLYAEHSLDLDP		
	SEQ ID NO: 315	1279 bp	
NOV87e, 308559628 DNA Sequence	GCCCTCAGAGGTAGGGGTGATTAGGGGTGTGCTCCATGATGGTCAGAAGCGCCAGCCAGCTCTC CTTGGCTGTGGGATGATCTGGGAGGCTGGCAGCAGGAAGCCATAGCGCCAGTGTCCCGGAG CTCGAAGGTGAAGGAGTACTTGATGCCCTGGCTGTAGGTCCAGTCAATAGTGTCTCCACTGGC TTGATAAATTGCCTTGATGTGTCATAGTTGAAGTGGTCCCGTAGAGAGAGCCAGGGC TGTCACAGCAGCCTTGGAAGCTGATCCAGCTCATCTGGTCAGGGACTGGTTCTGTTTGTGA GCCATAGGGATACATGAGGAGCTGGGAGTAGCTGTGGATGGAGATGAAGGCCTTGATGTTCCC ATGGTCCTTCAAAAGTCTACAATGGACTTGACCTCCACTTCGGAATTGGCAAACTTGCCGTG GTAAGTCTCCGAGCAGGGGTACTGTGGCTCCGGACAACCCAAAGCCAGCGTCCAGGTTCTCT GTTGGGGTCCAGCCCAATACAGAGGGAGCCTGCTGTGTGGGACGAGTCTTGCCGCACATGCG ATTCTGTCTGTGCGTGAAGGCAAGCCATCAGGGTTGGTGACGATCTCCAGGAAGATGTCCAA GGTGTGAGAAATGGCGGTGAAGCTGCATCCTGCCCGTAGTCTTGAGTGATCTTCTTTGCAAA CCAGACCCCACTGGCTGGGTGACCCACTCCCGGGAATGGATGCCCGTGTGATCCAGATGGC TGGACGCTTACTGCCCCCGTGTGAACCTCAGCACGTAAATGGGACGCCCTTCATAGGTGTT GCCAATCTGGATCTTGCTGACAAGGTGCGGGTTCTCCGCCACCAGCAGGTCCAGGAAGTCATA GATCTCTCCAGGGTGTGGTAGGTGGCGTAGTTAAAGTGTGGTGGAGCGGCCCGGGACCG GAAGCGAACATCTGCTCCTGCTCCTCGTCCAGCAGCGACTGCACGTCTCGATCATGGTCTC ATAGCTGATGCCGTGGGACTCCAGAAAGATCTTGACCGCTGGATGCTGGGGAAGGGCACTCG GACGTGATGGGGGAGGGTGGGAGGCCCGCCAGAAAGTCCAGTCCAGGTGCGAGGTCTCAG GTCTCCAGCTCCTTACCTTCTGTACCTGGGCCTCATCGGCTACAGAGATTCGGAGCACCTG ATGCCCCACAAAGTCTCCTTGCCAAAGACAGCCCCAACAGGACACTCAACACCAGCAACCC CCGCATGGTGGATCCGGTG		
	ORF Start: at 446		ORF Stop: TAG at 668

	SEQ ID NO: 316	74 aa	MW at 8100.6kD
NOV87e, 308559628 Protein Sequence	VSEQGLLLAPDNPKPASQFLLGSTPIQREPAVWDRVLRHMRFVLCVKAKPSGLVTISRKMSKV SRMAVKAASCP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 87B.

Table 87B. Comparison of NOV87a against NOV87b through NOV87e.		
Protein Sequence	NOV87a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV87b	1..214 1..214	213/214 (99%) 213/214 (99%)
NOV87c	1..214 5..218	213/214 (99%) 213/214 (99%)
NOV87d	1..214 1..214	214/214 (100%) 214/214 (100%)
NOV87e	No Significant Alignment Found.	

5

Further analysis of the NOV87a protein yielded the following properties shown in Table 87C.

Table 87C. Protein Sequence Properties NOV87a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3577 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

10

A search of the NOV87a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 87D.

Table 87D. Geneseq Results for NOV87a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV87a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB41812				e-121

	sequence SEQ ID NO:3152 - Homo sapiens, 214 aa. [WO200058473-A2, 05-OCT-2000]	1..214	214/214 (100%)	
ABB65103	Drosophila melanogaster polypeptide SEQ ID NO 22101 - Drosophila melanogaster, 482 aa. [WO200171042-A2, 27-SEP-2001]	8..214 279..482	63/209 (30%) 120/209 (57%)	6e-24
AAU79185	Human phosphatidylinositol-four-phosphate adaptor protein-2 (FAPP-2) - Homo sapiens, 507 aa. [WO200212276-A2, 14-FEB-2002]	23..177 326..472	40/156 (25%) 77/156 (48%)	6e-07
ABG20808	Novel human diagnostic protein #20799 - Homo sapiens, 391 aa. [WO200175067-A2, 11-OCT-2001]	23..177 210..356	40/156 (25%) 77/156 (48%)	6e-07
AAB95725	Human protein sequence SEQ ID NO:18600 - Homo sapiens, 193 aa. [EPI074617-A2, 07-FEB-2001]	23..177 12..158	39/156 (25%) 77/156 (49%)	8e-07

In a BLAST search of public sequence databases, the NOV87a protein was found to have homology to the proteins shown in the BLASTP data in Table 87E.

Table 87E. Public BLASTP Results for NOV87a				
Protein Accession Number	Protein/Organism/Length	NOV87a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q99LU9	Hypothetical 24.6 kDa protein - Mus musculus (Mouse), 216 aa.	1..214 1..216	169/216 (78%) 193/216 (89%)	9e-95
AAH30735	Similar to hypothetical protein, MGC:7473 - Mus musculus (Mouse), 321 aa.	17..213 119..320	74/204 (36%) 108/204 (52%)	1e-27
AAH25515	Hypothetical protein - Mus musculus (Mouse), 207 aa.	17..213 5..206	74/204 (36%) 108/204 (52%)	1e-27
AAM70862	CG30392-PA - Drosophila melanogaster (Fruit fly), 223 aa.	8..214 20..223	63/209 (30%) 120/209 (57%)	2e-23
Q9W214	CG10509 protein - Drosophila melanogaster (Fruit fly), 482 aa.	8..214 279..482	63/209 (30%) 120/209 (57%)	2e-23

5

PFam analysis predicts that the NOV87a protein contains the domains shown in the Table 87F.

Table 87F. Domain Analysis of NOV87a

Pfam Domain	NOV87a Match Region	Identities/ Similarities for the Matched Region	Expect Value

Example 88.

The NOV88 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 88A.

5

Table 88A. NOV88 Sequence Analysis			
	SEQ ID NO: 317	1911 bp	
NOV88a, CG59983-01 DNA Sequence	CGGCGAGTGAATCTACCCAGGAGCTGAAGCTTCCCTTGGCAACTTGGGGACACCAGGGCCACA GTGACTTTGTGGCTTGGCTTTAGATAAAATTGACCATGGGCTGTAGAACCAGCAGCTCAGAA TCCATTAAAAGGAGAGCTGGGAGGAGAATGAAGAAGATGAGCAACATTTATGAGTCCGCTGCC AACACACTGGGAATCTTTAACAGCCCCTGCCTGACCAAAGTTGAGCTGCGTGTGGCGTGCAAA GGCATTCTGACAGAGATGCCCTTTCAAACCAGGCCCCCTGTGTCATCCTCAAGATGCAGTCT CATGGGCAGTGGTTTGAGGTTGACAGGACTGAGGTGATTTCGCACCTGCATAAACCCAGTGTAC TCAAACTGTTTACTGTGGACTTTTACTTTGAGGAGGTGCAGCGCCTGCGGTTTGAAGTCCAT GACATCAGCAGCAACCACAATGGGCTGAAGGAGGCCGACTTCCTTGGTGGCATTGGAGTGCACA CTTGGCCAGATTGTTTCCAGAGAAAGCTGTCCAAATCCTTGCTGAAGCATGGGAACACAGCA GGGAAATCTTCCATCGCGGTGATTGCTGAAGAATTATCTGGCAATGACGACTATGTTGAGCTT GCATTCAATGCACGGAATTTGGATGACAAGGATTTCTTCAGTAAATCTGACCCATTTCTGGAA ATTTTTCGTATGAATGATGATGCAACTCAGCAGCTGGTGCACCGAACTGAGGTTGTGATGAAT AACTTAAGCCCGAGCTGGAAATCATTCAAAGTATCTGTAAATCTCTATGCAGCGGAGACCCA GACCGCCGGCTAAAGTGCATAGTATGGGACTGGGACTCCAATGGCAAGCATGACTTCATTGGA GAATTCACCTCGACATTCAAGGAGATGAGAGGAGCAATGGAAGGGAAACAGCTGCGAGTGGAG TGCATCAATCCCAAGTACAAAGCCAAGAAGAATTACAAGAACTCAGGCACTGTGATTCTG AATCTGTGCAAGATTACAAGATGCATTCTTCTTGGACTACATCATGGGTGGCTGCCAAATC CAGTTTACAGTAGCTATAGATTTCACTGCCTCAAACGGGGACCCAGGAACAGCTGTTCCTTG CACTACATCCACCCTTACCAACCAATGAGTATCTGAAAGCTTTGGTAGCTGTGGGGGAGATT TGCCAAGACTATGACAGTGCACAAATGTTCCCTGCCTTGGGTTTGGGCCAGGATACCTCCA GAGTACACGGTCTCTCATGACTTTGCAATCACTTTAATGAAGAACAACCCAAATGTGCAGGA ATTCAAGGAGTTGTGGAAGCCTATCAGAGCTGTCTTCTTAAGCTCCAACCTTACGGTCCCACC AACATTGCCCCCATCATCCAGAAGGTGCGCAAGTCAGCGTCAGAGGAACTAACACCAAGGAG GCATCGCAATACTTCATCTGCTGATCCTGACAGATGGTGTATCAGAGACATGGCCGACACC CGGGAGGCCATTGTCCATGCCTCCACCTCCCATGTGAGTCATCATCGTGGGAGTAGGGAAC GCTGACTTCAGTGACATGAGATGCTGGACGGTGATGATGGGATTCTGAGGTACCCAGGGA GAGCCTGTTCTTCGAGACATCGTCCAGTTCGTGCCCTTCAGGAACCTCAAACACGCATCTCCA GCTGCCCTGGCAAAGAGCGTGTGGCTGAAGTCCCAACCAAGTTGTGGACTATTACAATGGC AAAGGAATTAACCAAAATGTTTCATCAGAAATGTATGAATCTTCAGCACACTAGCACCATGA ACTCCCCACACAGTTTACAGAGTTCTGAAATACTATTCTGCTAATATTTTCATATTTAATAC TTCTACTATTCTGCAAATGG		
	ORF Start: ATG at 154		ORF Stop: TGA at 1825
	SEQ ID NO: 318	557 aa	MW at 62264.4kD
NOV88a, CG59983-01 Protein Sequence	MKMSNIYESAANTLGI FNSPLTKVELRVACKGISDRDLSKPGPCVILKMQSHGQWFEVDR TEVIRTCINPVYSKLFTVDFYFEVQRLRFVHDISSNHNLKEADFLGGMECTLGQIVSQRK LSKSLKKGNTAGKSSIAVIAEELSGNDDYVELAFNARKLDDKDFFSKSDPFLEIFRMNDDAT QQLVHRTEVVMNLSPAWKSFKVSVNSLCSGDPDRRLKCI VWDWDSNGKHDFIGFTSTFKEM RGAMEGKQVQWECINPKYKAKKKNYKNSGTVILNLCKIHKMHSFLDYIMGGCQIQFTVAIDFT ASNGDPRNSCSLHYIHPYQPNELKALVAVGEICQDYSDKMFPAFGFGARIPPEYTVSHDFA INFNEEQPKCAGIQGVVEAYQSLPKLQLYGPTNIAPIIQKVAKSASEETNTKEASQYFILLI LTDGVIDTMDATREAI VHASHLPMSVII VGVGNADFSDMQMLDGGDGLRSPKGEPEVLRDIVQ FVPPRNFKHASPAALAKSVLAEPVNVQVVDYNGGKIKPKCSSEMYESSSTLAP		
	SEQ ID NO: 319	1795 bp	
NOV88b,	AGTGACTTTGTGGCTTGGCTTTAGATAAAATTGACCATGGCTGTAGAACCAGCAGCTCAGAA		

CG59983-02 DNA Sequence	TCCATTAAAAGGAGAGCTGGGAGGAGAATGAAGAAGATGAGCAACATTTATGAGTCCGCTGCC AACACACTGGGAATCTTTAACAGCCCTGCCTGACCAAAGTTGAGCTGCGTGTGGCGTGCAAA GGCATTCTGACAGAGATGCCCTTTCCAAACCAGACCCCTGTGTCTATCCACAAGATGCGAGTCT CATGGGCAGTGGTTTGAGGTTGACAGGACTGAGGTGATTGCGACCTGCATAAACCCAGTGATC TCAAAACTGTTTACTGTGGACTTTTACTTTGAGGAGGTGAGCGCCTGCGGTTTGAAGTCCAT GACATCAGCAGCAACCACAATGGGCTGAAGGAGGCCGACTTCCTTGGTGGCATGGAGTGCACA CTTGGCCAGATTGTTTCCAGAGAAAGCTGTCCAAATCCTTGCTGAAGCATGGGAACACAGCA GGGAAATCTTCCATCACGGTGATTGCTGAAGAATTATCTGGCAATGACGACTATGTTGAGCTT GCATTCAATGCACGGAATGGATGACAAGGATTCTTCAGTAAATCTGACCCATTCTTGGA ATTTTTCGTATGAATGATGATGCAACTCAGCAGCTGGTGCACCGAACTGAGGTTGTGATGAAT AACTTAAGCCAGCCTGGAAATCATTCAAAGTATCTGTAAATTCTCTATGCAGCGGAGACCCA GACCGCCGGCTAAAGTGATAGTATGGGACTGGGACTCCAATGGCAAGCATGACTTCATTGGA GAATTCACCTCGACATTCAAGGAGATGAGAGGAGCAATGGAAGGGAACAGGTGCAGTGGGAG TGCATCAATCCCAAGTACAAAGCCAAGAAGAATTACAAGAATCAGGCACTGTGATTCTG AATCTGTGCAAGATTACAAGATGCATTCTTCTTGGACTACATCATGGGTGGCTGCCAAATC CAGTTTACAGTAGCTATAGATTTCACTGCCTCAAACGGGGACCCAGGAACAGCTGTTCCTTG CACTACATCCACCCTTACCAACCAATGAGTATCTGAAGCTTTGGTAGCTGTGGGGGAGATT TGCCAAGACTATGACAGTGACAAAATGTTCCCTGCCTTTGGGTTTGGCGCCAGGATACCTCCA GAGTACACGGTCTCTCATGACTTTGCAATCAACTTTAATGAAGACAACCCAGAATGTGCAGGA ATTCAAGGAGTTGTGGAAGCCTATCAGAGCTGTCTTCTAAGCTCCAACCTACGGTCCCAACC AACATTGCCCCCATCATCCAGAAGGTTGCCAAGTCGGCGTCAGAGGAAACTAACACCAAGGAG GCATCGCAATACTTTCATCTGCTGATCCTGACAGATGGTGTATCACAGACATGGCCGACACC CGGGAGGCCATTGTCCATGCCTCCACCTCCCATGTCTAGTCATCATCGTGGGAGTAGGGAAC GCTGACTTCAGTGACATGCAGATGCTGGACGGTGATGATGGGATTCTGAGGTACCCAAGGGA GAGCCTGTTCTTCGAGACATCGTCCAGTTCGTGCCCTTCAGGAACCTCAAACATGCATCTCCA GCTGCCCTGGCAAAGAGCGTGTGGCTGAAGTCCCAAACCAAGTTGTGGACTATTACAATGGC AAAGGAATTAACCAAAATGTTTCATCAGAAATGTATGAATCTTCCAGAACACTAGCACATGA ACTCCCAACACAGTTTACAGAGTTCTGAAA		
	ORF Start: ATG at 91		ORF Stop: TGA at 1762
	SEQ ID NO: 320	557 aa	MW at 62418.4kD
NOV88b, CG59983-02 Protein Sequence	MKKMSNIYESAANTLGFINSPLTKVELRVACKGISDRDALSKPDPCVIHKMQSHGQWFEVDR TEVIRTCINPVYSKLFVDFYFEEVQRLRFVHDISSNHNLKEADFLGGMECTLGQIVSQRK LSKSLKKGNTAGKSSITVIAEELSGNDDYVELAFNARKLDDKDFFSKSDPFLEIFRMNDDAT QQLVHRTEVVMNLSPAWKSFKVSVNSLCSGDPDRRLKCIWVDWDSNGKHDFIGFTSTFKEM RGAMEGKQVQWECINPKYKAKKNYKNSGTVILNLCKIHKMHSFLDYIMGGCQIQFTVAIDFT ASNGDPRNSCSLHYIHPYQNEYLKALVAVGEICQDYDSDKMFPAPFGFARIPPEYTVSHDFA INFNEDNPECAGIQGVVEAYQSLPKLQLYGPTNIAPIIQKVAKSASEETNTKEASQYFILLI LTDGVIITDMADTREAIIVHASHLPMSVIIVGVGNAFSDQMQLDGGDGLRSPKGEPLRDIQV FVPRFRFKHASPAALAKSVLAEPVNPQVVDYNGKGIKPKCSSEMYESSRTLAP		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 88B.

Table 88B. Comparison of NOV88a against NOV88b.		
Protein Sequence	NOV88a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV88b	1..557 1..557	529/557 (94%) 531/557 (94%)

5

Further analysis of the NOV88a protein yielded the following properties shown in Table 88C.

Table 88C. Protein Sequence Properties NOV88a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV88a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 88D.

5

Table 88D. Geneseq Results for NOV88a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV88a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
ABB06047	Human NS protein sequence SEQ ID NO:139 - Homo sapiens, 564 aa. [WO200206315-A2, 24-JAN-2002]	11..543 9..543	375/535 (70%) 453/535 (84%)	0.0
ABB10990	Human copine VII homologue, SEQ ID NO:1360 - Homo sapiens, 415 aa. [WO200157188-A2, 09-AUG-2001]	60..444 8..394	354/387 (91%) 362/387 (93%)	0.0
AA Y49835	Mouse neuronal activity regulated C2-domain containing protein - Mus sp, 557 aa. [JP11269198-A, 05-OCT-1999]	3..542 5..541	340/543 (62%) 429/543 (78%)	0.0
AA Y49836	Human neuronal activity regulated C2-domain containing protein - Homo sapiens, 557 aa. [JP11269198-A, 05-OCT-1999]	24..542 20..541	334/522 (63%) 424/522 (80%)	0.0
AA Y49834	Mammalian brain specific generic protein - Mammalia, 557 aa. [JP11269198-A, 05-OCT-1999]	24..542 20..541	333/522 (63%) 419/522 (79%)	0.0

In a BLAST search of public sequence databases, the NOV88a protein was found to have homology to the proteins shown in the BLASTP data in Table 88E.

Table 88E. Public BLASTP Results for NOV88a				
Protein Accession Number	Protein/Organism/Length	NOV88a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value

Q8TEX1	Copine-like protein isoform B - Homo sapiens (Human), 575 aa.	1..557 19..575	551/557 (98%) 553/557 (98%)	0.0
Q96A23	CDNA FLJ31613 fis, clone NT2RI2002958, moderately similar to Homo sapiens copine VI protein (Similar to RIKEN cDNA 3632411M23 gene) (Copine-like protein isoform A) - Homo sapiens (Human), 557 aa.	1..557 1..557	551/557 (98%) 553/557 (98%)	0.0
Q9Z140	Copine VI (Neuronal-copine) (N-copine) - Mus musculus (Mouse), 557 aa.	3..542 5..541	340/543 (62%) 429/543 (78%)	0.0
O95741	Copine VI (Neuronal-copine) (N-copine) - Homo sapiens (Human), 557 aa.	24..542 20..541	334/522 (63%) 424/522 (80%)	0.0
Q8WVG1	Copine VI (neuronal) - Homo sapiens (Human), 557 aa.	24..542 20..541	333/522 (63%) 424/522 (80%)	0.0

PFam analysis predicts that the NOV88a protein contains the domains shown in the Table 88F.

Table 88F. Domain Analysis of NOV88a			
Pfam Domain	NOV88a Match Region	Identities/ Similarities for the Matched Region	Expect Value
C2	28..115	20/100 (20%) 59/100 (59%)	0.085
C2	161..246	26/102 (25%) 62/102 (61%)	0.00057

5

Example 89.

The NOV89 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 89A.

Table 89A. NOV89 Sequence Analysis			
	SEQ ID NO: 321	724 bp	
NOV89a, CG93335-01 DNA Sequence	CAGGAGGCGGGTGGGTCAAGGTAAGTCTGGGCTACAGAGTCCTTGCTGGGGGTTCCGGGAGCG CTTGGACCCCGGCTTCTGGGACGCGTCAGAATATTATCCAGCAATGCAAATGAACAACTATA ACTACACACAGCTGCATGGATAAATGTCAGAAACATGACGTTGAGTGTGAGAAGCCAGATGCA AACGAGGACTCACTGTGCAATTCTGTGCATGTACAGTGCCAGGAGAAGGGAGCACTGGCTTT GCTTTCATCAGGCCAAAGATGCCCTTTCTTTGGGAATACGTTTCAGTCCGAAGAAGACACCTCCT CGGAAGTCGGCATCTCTCCAACCTGCATTCTTTGGATCGATCAACCCGGGAGGTGGAGCTG GGCTTGAATACGGATCCCCGACTATGAACCTGGCAGGGCAAAGCCTGAAGTTTGAAATGGC CAGTGGATAGCAGAGACAGGGTTAGTGGCGGTGTGGACCGGAGGGAGGTTTCAGCGCCTTCGC AGGCGGAACCAGCAGTTGGAGGAAGAGAACAATCTCTTGGCGGCTGAAAGTGGACATCTTATTA GACATGCTTTCAGAGTCCACTGCTGAATCCCACTTAATGGAGAAGGAAGTGGATGAAGTGGAG ATCAGCCGGAAGAGAAAAATGAAGACCCAGAGACATTTATTGGGGAGTAGGATGTGGCTGAGT		

	GCTTTTTTTTGGCCAGACTAGCGGATTGAG		
	ORF Start: ATG at 142		ORF Stop: TGA at 649
	SEQ ID NO: 322	169 aa	MW at 19286.6kD
NOV89a, CG93335-01 Protein Sequence	MDKCQKHDVECEKPDANEDSLCNSVHVQWPGEGSTGFAFIRPKMPFFGNTFSPKKTTPRKSAS LSNLHSLDRSTREVELGLEYGSPMTNLAGQSLKFENGQWIAETGVSGGVDRREVQRLRRRNQ LEEENLLRLKVDILLDMLSESTAESHLMKELDELRI SRKRK		
	SEQ ID NO: 323	615 bp	
NOV89b, CG93335-02 DNA Sequence	CGTCAGAATATTATCCAGCAATGCAAATGAACAACTATAACTACACACAGCTGCATGGATAA ATGTCAGAAACATGACGTTGAGTGTGAGAAGCCAGATGCAAACGAGGACTCACTGTGCAATTC TGTGCATGTACAGTGGCCAGGAGAAGGGAGCACTGGCTTTGCTTTCATCAGGCCAAAGATGCC TTTCTTTGGGAATACGTTTCAGTCCGAAGAAGACACCTCCTCGGAAGTCGGCATCTCTCCAA CCTGCATTCTTTGGATCGATCAACCGGGAGGTGGAGCTGGGCTTGAATACGGATCCCGAC TATGAACCTGGCAGGGCAAAGCCTGAAGTTTGAAAAATGGCCAGTGGATAGCAGAGACAGGGGT TAGTGGCGGTGTGGACCGGAGGGAGGTTTCAGCGCCTTCGAGGCGGAACAGCAGTTGGAGGA AGAGAACAAATCTCTTCGGCTGAAAGTGGACATCTTATTAGACATGCTTTTCAGAGTCCACTGC TGAATCCCACTTAATGGAGAAGGAAGTGGATGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAG ACCCAGAGACATTTATTGGGGAGTAGGATGTGGCTGAGTGCTTTTTT		
	ORF Start: ATG at 56		ORF Stop: TGA at 563
	SEQ ID NO: 324	169 aa	MW at 19286.6kD
NOV89b, CG93335-02 Protein Sequence	MDKCQKHDVECEKPDANEDSLCNSVHVQWPGEGSTGFAFIRPKMPFFGNTFSPKKTTPRKSAS LSNLHSLDRSTREVELGLEYGSPMTNLAGQSLKFENGQWIAETGVSGGVDRREVQRLRRRNQ LEEENLLRLKVDILLDMLSESTAESHLMKELDELRI SRKRK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 89B.

Table 89B. Comparison of NOV89a against NOV89b.		
Protein Sequence	NOV89a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV89b	1..169 1..169	146/169 (86%) 146/169 (86%)

5

Further analysis of the NOV89a protein yielded the following properties shown in Table 89C.

Table 89C. Protein Sequence Properties NOV89a	
PSort analysis:	0.4600 probability located in nucleus; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV89a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 89D.

Table 89D. Geneseq Results for NOV89a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV89a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAM00955	Human bone marrow protein, SEQ ID NO: 431 - Homo sapiens, 175 aa. [WO200153453-A2, 26-JUL-2001]	31..169 37..175	139/139 (100%) 139/139 (100%)	1e-74
AA Y86201	Nuclear transport protein clone hfb2025 protein sequence - Homo sapiens, 67 aa. [WO9964455-A1, 16-DEC-1999]	103..169 1..67	67/67 (100%) 67/67 (100%)	6e-30
ABB23535	Protein #5534 encoded by probe for measuring heart cell gene expression - Homo sapiens, 26 aa. [WO200157274-A2, 09-AUG-2001]	44..69 1..26	26/26 (100%) 26/26 (100%)	2e-08
AAB69070	Human male enhanced antigen-2 (MEA-2) protein sequence SEQ ID NO:2 - Homo sapiens, 1374 aa. [JP2000316580-A, 21-NOV-2000]	62..163 768..868	25/102 (24%) 45/102 (43%)	1.2
ABB06335	Human GDMPLP-1 orthologue BAA93660 protein sequence - Homo sapiens, 1694 aa. [WO200192524-A2, 06-DEC-2001]	116..151 917..952	16/36 (44%) 25/36 (69%)	1.6

5

In a BLAST search of public sequence databases, the NOV89a protein was found to have homology to the proteins shown in the BLASTP data in Table 89E.

Table 89E. Public BLASTP Results for NOV89a				
Protein Accession Number	Protein/Organism/Length	NOV89a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
Q9Y3M2	Protein C22orf2 (Cytosolic leucine-rich protein) (HRIHFB2025) - Homo sapiens (Human), 126 aa.	44..169 1..126	126/126 (100%) 126/126 (100%)	4e-66
AAM73678	Leucine-rich protein - Bos taurus (Bovine), 127 aa.	44..169 1..127	115/127 (90%) 124/127 (97%)	1e-60

AAM73679	Leucine-rich protein - Rattus norvegicus (Rat), 127 aa.	44..169 1..126	105/126 (83%) 120/126 (94%)	1e-56
Q9DIC2	Protein C22orf2 homolog (Cytosolic leucine-rich protein) - Mus musculus (Mouse), 127 aa.	44..169 1..126	104/126 (82%) 120/126 (94%)	1e-56
AAM73681	Leucine-rich protein - Brachydanio rerio (Zebrafish) (Danio rerio), 125 aa.	44..169 1..124	93/126 (73%) 114/126 (89%)	2e-49

PFam analysis predicts that the NOV89a protein contains the domains shown in the Table 89F.

Table 89F. Domain Analysis of NOV89a			
Pfam Domain	NOV89a Match Region	Identities/ Similarities for the Matched Region	Expect Value

5

Example 90.

The NOV90 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 90A.

Table 90A. NOV90 Sequence Analysis			
	SEQ ID NO: 325	10677 bp	
NOV90a, CG94377-01 DNA Sequence	GGATCGAATCGCGGCCGCTCGACGGTTGACAGGCGCTTCCCTCTGGAAGTGGCGACTGCTG CGGGCCTGAGCGCTGGTCTCACGCGCTCGGGAGCCAGGTGGCGGCCGATGAGGCGCAGCA AGGCTGACGTGGAGCGGTACATCGCCTCGGTGCAGGGCTCCACCCCGTCGCCTCGACAGAAGT CAATGAAAGGATTCTATTTGCAAAGCTGTATTATGAAGCTAAAGAATATGATCTTGCTAAAA AATACATATGTACTTACATTAATGTGCAAGAGAGGGATCCCAAAGCTCACAGATTTCTGGGTC TTCTTTATGAATTGGAAGAAAACACAGACAAAGCCGTTGAATGTTACAGGCGTTCACTGGAAT TAAACCAACACAAAAGATCTGTGTTGAAGATTGCAGAATTGCTTTGTAAAAATGATGTTA CTGATGGAAGAGCAAAATACTGGCTTGAAGAGCAGCCAACTTTTCCAGGAAGTCTTGCAA TTTATAAACTAAAGGAACAGCTTCTAGATTGTGAAGGTGAAGATGGATGGAATAAACTTTTG ACTTGATTCAGTCAGAACTTTATGTAAGACCTGATGACGTCCATGTGAACATCCGGCTAGTGG AGGTGTATCGTCACTAAAAGATTGAAGGATGCTGTGGCCCACTGCCATGAGGCAGAGAGGA ACATAGCTTTGCGTTCAAGTTTAGAATGGAATTCGTGTGTTGTACAGACCTTAAGGAATATC TGGAGTCTTTACAGTGTGAGGCTGTGATAAAGTGACTGGCGAGCAACCAATACAGACTTAC TGCTGGCCTATGCTAATCTTATGCTTCTTACGCTTTCCTACTAGAGATGTGCAGGAAAGTAGAG AATTACTGCAAAGTTTGTAGTGTCTTCACTGTGTGAAATCTTTGGGTGGAATGATGAAC TGTCAGCTACTTTCTAGAAATGAAAGGACATTCTACATGCATGCTGGTCTCTGCTTTTGA AGATGGGTCAGCATAGTAGTAATGTTCAATGGCGAGCTCTTTCTGAGCTGGCTGCATTGTGCT ATCTCATAGCATTTCAGGTTCCAAGACCAAGATTAAATTAATAAAAGGTGAAGCTGGACAAA ATCTGCTGGAATGATGGCCTGTGACCGACTGAGCCAATCAGGGCACATGTTGCTAAACTTAA GTCGTGGCAAGCAAGATTTTTTAAAGAGATTGTTGAACTTTTGCCAAACAAAGCGGGCAGT CTGCATTATATGATGCTCTGTTTTCTAGTCAGTCACCTAAGGATACATCTTTCTTGGTAGCG ATGATATTGGAACATTGATGTACGAGAACCAGAGCTTGAAGATTGACTAGATACGATGTTG GTGCTATTTCGAGCACATAATGGTAGTCTTCAGCACCTTACTTGGCTTGGCTTACAGTGAATT CATTGCTGCTTTACCTGGAATCCGAAAATGGCTAAAACAGCTTTTCCATCATTGCCCCATG AAACCTCAAGGCTTGAACAAATGCACCTGAATCAATATGTATTTAGATCTTGAAGTATTC		

TCCTTGGAGTAGTATATACCAGCCACTTACAATTAAGGAGAAATGTAATTCTCACCACAGCT
CCTATCAGCCGTTATGCCTGCCCTTCCTGTGTGTAACAGCTTTGTACAGAAAGACAAAAAT
CTTGGTGGGATGCGGTTTGTACTCTGATTACAGAAAAGCAGTACCTGGAAACGTAGCAAAAT
TGAGACTTCTAGTTCAGCATGAAATAAACACTCTAAGAGCCCAGGAAAAACATGGCCCTTCAAC
CTGCTCTGCTTGTACATTGGGCAGAAATGCCTTCAGAAAAACGGGCAGCGTCTTAATCTTTTT
ATGATCAACGAGAATACATAGGGAGAAGTGTTCATTATTGGAAGAAAGTTTTGCCATTGTGTA
AGATAATAAAAAAGAAAGACAGTATTCCTGAACCTATTGATCCTCTGTTTAAACATTTTCATA
GTGTAGACATTACGGCATCAGAAATTGTTGAATATGAAGAAGACGCACATAAATTTTGCTA
TATTGGATGCAGTAAATGGAAATATAGAAGATGCTGTGACTGCTTTTGAATCTATAAAAAAGTG
TTGTTTCTTATTGGAATCTTGACTGATTTTTACAGGAAGGCAGAAACATTGAAAATGATG
CCCTTTCTCCTGAAGAACAAGAAGATGCAAAAATTATCTGAGAAAGACCAGGGACTACCTAA
TAAAGATTATAGATGACAGTGATTCAAATCTTTCAGTGGTCAAGAAATTGCCTGTGCCCTGG
AGTCTGTAAAAGAGATGCTTAATTCAGTCATGCAGGAAGCTGAAGACTATAGTGAAGGAGGTC
CTCTCTATAAAAAATGGTTCTTTGGCAAAATGCAGATTGAGAAATAAAACGTTCTACACCGTCTC
CTACCAGATATTCATATCACCAGTAAAGTTACAAGTATTCTCCCAAAACACCACCTCGAT
GGGCAGAAAGATCAGAAATCTTTACTGAAAATGATTGGCCAACAAGTAGAGGCCATTAAAGAAAG
AAATGCAGGAGTTGAAACTAAATAGCAGTAACTCAGCATCCCTCATCGTTGGCCCCACAGAGA
ATTATGGACCAGACTCGGTGCCTGATGGATATCAGGGGTACAGACATTTTCATGGGGTCCAC
TAACAGTTGCAACTACTGGCCCTTCAGTATATTATAGTCAGTCACCAGCATATAATTTCCAGT
ATCTTCTCAGACCAGCAGCTAATGTTACTCCCAACAAGGGCCAGTCTATGGCATGAATAGGC
TTCCACCCCAACAGCATATTTATGCCTATCCGCAACAGATGCACACACCGCCAGTGCAAGCT
CATCTGCTGTATGTTCTCTCAGGAGATGTATGGTCTCTGCATTGCGTTTTGAGTCTCCTG
CAACGGGAATTTCTATCGCCAGGGGTGATGATTACTTTAATTACAATGTTCAACAGACAAGCA
CAAATCCACCTTTGCCAGAACCAGGATATTTCAAAAACCTCCGATTGCAGCTCATGCTTCAA
GATCTGCAGAATCTAAGACTATAGAATTTGGGAAAATAATTTTGTTCAGCCCATGCCGGGTG
AAGGATTAAGGCCATCTTTGCCAACACAAGCACACAACACAGCCAACCTCTTTTAAATTTA
ACTCAAATTTCAAATCAAATGATGGTGACTTCACGTTTTCTCACCACAGTTGTGACACAGC
CCCCTCTGCAGCTTACAGTAACAGTGAAAGCCTTTTAGGTCTCTGACTTCAGATAAACCTT
TGCAAGGAGATGGCTATAGTGGAGCCAAACCAATTCCTGGTGGTCAAACCATTTGGGCCCTCGAA
ATACATTCAATTTTGAAGCAAAAATGTGTCTGGAATTTCAATTTACAGAAAACATGGGTCTGA
GTCAGCAAAAGAAATCTGGTTTTTCGGCGAAGTGATGATGTTTACTTTCCAGGTCAGGGA
AATCAGTATTTGGAACACCCACTTTAGAGACAGCAAAACAAGAAATCATGAGACAGATGGAGGAA
GTGCCCATGGGATGATGATGATGACGGTCTCACTTTGAGCCTGTAGTACCTCTTCTCTGATA
AGATTGAAGTAAAACTGGTGAGGAAGATGAAGAAGAAATCTTTTGCAACCGCGCGAAATTTGT
TTCGTTTCGATGTAGAATCCAAAGAAATGGAAAGAACGTGGGATTGGCAATGTAAAAATACTGA
GGCATAAAACATCTGGTAAATTCGCCTTCTAATGAGACGAGAGCAAGTATTGAAAATCTGTG
CAAATCATTACATCAGTCCAGATATGAAATTGACACCAAATGCTGGATCAGACAGATCTTTTG
TATGGCATGCCCTTGATTATGCAGATGAGTTGCCAAAACAGAACTTGTCTATTAGGTTCA
AACTCCTGAGGAAGCAGCACTTTTAAATGCAAGTTTGAAGAAGCCAGAGCATTTTAAAG
CCCCAGGAACAAATGTAGCCATGGCGTCAAATCAGGCTGTGAGAAATGTAAAAGAACCCACAA
GTCATGATAACAAGGATATTTGCAAATCTGATGCTGGAACCTGAATTTTGAATTTCAGGTTG
CAAAGAAAGAAGGCTCTTGGTGGCATTGTAACAGCTGCTCATTAAAGAATGCTTCAACTGCTA
AGAAATGTGTATCATGCCAAATCTAAACCAAGCAATAAGAGCTCGTTGGCCACCATTAG
CTGAACTGTTTTACTCCTAAAACAGCCAGAGAATGTTCAAGATCGATTGCTATTGGTGA
CTCAAAGAAAGAAGGTCAGTGGGATTGTAGTATTGTTTAGTAAGAAATGAACCTACTGTAT
CTAGGTGCATTGCGTGTGAGAATACAAATCTGCTAACAAAAGTGGATCTTCATTTGTTTCATC
AAGCTTCATTTAAATTTGGCCAGGAGATCTTCTTAAACCTATTAAACAGTGATTTTCAGATCTG
TTTTTTCTACAAAGGAAGGACAGTGGGATTGCAGTGATGTTTGGTACAAAATGAGGGGAGCT
CTACAAAATGTGCTGCTTGTGAGAAATCCGAGAAAACAGAGTCTACCTGCTACTTCTATTCCAA
CACCTGCCTCTTTAAGTTTGGTACTTCAGAGACAAGTAAACTCTAAAAAGTGGATTGAAAG
ACATGTTTGCTAAGAAGGAAGGACAGTGGGATTGCAGTTCATGCTTAGTGCGAAATGAAGCAA
ATGCTACAAGATGTGTTGCTTGTGAGAATCCGGATAAACCAAGTCCATCTACTTCTGTTCCAG
CTCCTGCCTCTTTAAGTTTGGTACTTCAGAGACAAGCAAGGCTCCAAAGAGCGGATTGAGGGA
GAATGTTCACTAAGAAGGAGGACAGTGGGATTGCAGTGTGCTTGTAGTAAGAAATGAAGCCA
GTGCTACCAAATGTATTGCTTGTGAGAATCCAGGTAAACAAAATCAAATCTTCTGCAGTTT
CAACACCTGCCTCTTCAGAGACAAGCAAGGCTCCAAAGAGCGGATTGAGGGAATGTTCACTA
AGAAGGAGGACAGTGGGATTGCAGTGTGCTTAGTAAGAAATGAAGCCAGTGTACCAAAT
GTATTGCTTGTGAGAATCCAGTAAACAAAATCAAATCTTCTGCAGTTTCAACACCTGCCT
CTTCAGAGACAAGCAAGGCTCCAAAGAGCGGATTGAGGGAATGTTCACTAAGAAGGAAGGAC
AGTGGGATTGCAGTGTGCTTAGTAAGAAATGAAGCCAGTGTACCAAATGTATTGCTTGTG
AGTGTCCAAGTAAACAAAATCAAACACTGCAATTTCAACACCTGCCTCTTCGGAGATAAGCA
AGGCTCCAAAGAGTGGATTGAAAGGAATGTTTCATCAGGAAAGGACAGTGGGATTGTAGTGT
GCTGTGTACAAAATGAGAGTTCTTCTTAAATGTGTGGCTTGTGATGCTTAAACCACTC
ATAAACCTATTGCAGAAGCTCCTTCAGCTTTCACACTGGGCTCAGAAATGAAGTTGCATGACT
CTTCTGGAAGTCAAGTGGGAACAGGATTAAAAAGTAATTTCTCAGAAAAAGCTTCAAGTTTG
GCAATACAGAGCAAGGATTCAAATTTGGGCATGTGGATCAAGAAATTCACCTTCATTTATGT

TTCAAGGGTTCTTCTAATACAGAATTTAAGTCAACCAAAGAAGGATTTTCCATCCCTGTGTCTG
CTGATGGATTTAAATTTGGCATTTCGGAACCCAGGAAATCAAGAAAAGAAAAGTGAAAAGCCTC
TTGAAAATGGTACTGGCTTCCAGGCTCAGGATATTAGTGGCCAGAAGAAATGGCCGTGGTGTGA
TTTTTGGCCAAACAAGTAGCACTTTTACATTTGCAGATCTTGCAAAATCAACTTCAGGAGAAG
GATTTAGTTTGGCAAAAAGACCCCAATTTCAAGGGATTTTCAAGGTGCTGGAGAAAATTAT
TCTCATCACAAATACGGTAAAATGGCCAATAAAGCAAACACTTCCGGTGACTTTGAGAAAAGATG
ATGATGCCCTATAAGACTGAGGACAGCGATGACATCCATTTTGAACCAAGTAGTTCAAATGCCCG
AAAAAGTAGAACTTGTAACAGGAGAAGAAGATGAAAAAGTTCTGTATTCACAGCGGTAAAAC
TATTTAGATTTGATGCTGAGGTAAAGTCAGTGGAAGAAAAGGGGCTTGGGGAAGCTTAAAAATTC
TCAAAAACGAGGTCAATGGCAAACTAAGAATGCTGATGCGAAGAGAACAAGTACTAAAAGTGT
GTGCTAATCATTGGATAACGACTACGATGAACCTGAAGCCTCTCTCTGGATCAGATAGAGCAT
GGATGTGGTTAGCCAGTGATTTCTCTGATGGTGATGCCAACTAGAGCAGTTGGCAGCAAAAT
TTAAAACACCAGAGCTGGCTGAAGAATTCAAGCAGAAATTTGAGGAATGCCAGCGGCTTCTGT
TAGACATACCACTTCAAACCTCCCATAAACTTGTAGATACTGGCAGAGCTGCCAAGTTAATAC
AGAGAGCTGAAGAAATGAAGAGTGGACTGAAAGATTTCAAACATTTTGAACAAATGATCAAA
CAAAAGTCACTGAGGAAGAAAATAAGGGTTCAGGTACAGGTGCGGCCGGTGCTCAGACACAA
CAATAAAACCCAATCTGAAAACACTGGGCCACATTAGAATGGGATAACTATGATTTAAGGG
AAGATGCTTTGGATGATAGTGTAGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGTAGTGT
CTGTGAGAAAAATCTTTTCCGTTTGGTGAGTCAACAACAGGATTTAACTTCAGTTTAAAT
CTGCTTTGAGTCCATCTAAGTCTCTGCAAGTTGAATCAGAGTGGGACTTCAGTTGGCAGTGT
ATGAAGAATCTGATGTTACTCAAGAAGAAGAGAGAGATGGACAGTACTTTGAACCTGTTGTTT
CTTTACCTGATCTAGTTGAAGTATCCAGTGGTGAGGAAAATGAACAAGTTGTTTAACTCACA
GGGCAAAACTCTACAGATATGATAAAGATGTTGGTCAATGGAAAGAAAGGGCATTGGTGATA
TAAAGATTTTACAGAATTATGATAAAGCAAGTTCGTATAGTGATGAGAAGGGACCAAGTAT
TAAACTTTGTGCCAATCACAGAATAACTCCAGACATGACTTTGCAAAATATGAAAGGGACAG
AAAGAGTATGGTTGTGGACTGCTGTGATTTTGCAGATGGAGAAAAGAAAGTAGAGCATTTAG
CTGTTCTGTTTAAACTACAGGATGTTTGCAGACTCGTTTAAAGAAAATTTTGTATGAAGCAAAA
CAGCCAGGAAAAAGATTTCTTGATAACACCTCATGTTTCTCGGTCAAGCACTCCCAGAGAGT
CACCATGTGGCAAAATTTGCTGTAGCTGTATTAGAAGAAACCAAGAGAGAGGACAGATGTTA
TTCAGGGTGATGATGTAGCAGATGCAACTTCAGAAGTTGAAGTGTCTAGCACATCTGAAACAA
CACCAAAAGCAGTGGTTTCTCCTCCAAAGTTTGTATTGTTGTTTCAAGTCTGTAAAAGCATTT
TTAGTAGTGA AAAATCAAACCATTTGCATTCCGGCAACAGTTCAGCCACTGGGTCTTTGTTT
GATTTAGTTTTAATGCACCTTTGAAAAGTAACAATAGTGAAGTGTCTAGTAGCCAGAGTG
GATCTGAAAGCAAAGTGGAACTTAAAAATGTGAAGTGTCAAAGAACTCTGATATCGAACAGT
CTTCAGATAGCAAAGTCAAATCTCTTTGCTTCTTTCCAAACGGAAGAATCTCAATCACT
ACACATTTAAACACCAGAAAAGGCAAAAGAGAAGAAAACCTGAAGATTTCTCCCTCAGATG
ATGATGTTCTCATTGTATATGAACCTCAACCGCTGAGCAGAAAGCCCTTGCAACCAAAC
TTAACTTCTCCTCAACTTCTCTGCTACAAGAAATAGACCAGATTATGTTAGTGAAGAAGAGG
AGGATGATGAAGATTTGAAACAGCTGTCAAGAACTTAATGGAAAATATATTTGGATGGCT
CAGAAAAATGTAGACCTTTGGAAGAAAATACAGCAGATAATGAGAAAGAATGTATTATTTGTT
GGGAAAAGAAACCAACAGTTGAAGAGAAGGCAAAAGCAGATACGTTAAAACCTTCCACCTACAT
TTTTTTGTGGAGTCTGTAGTGATCTGATGAAGACAATGGAAATGGGGAAGACTTTCAATCAG
AGCTTCAAAGTTCAGGAAGCTCAAATCTCAGACAGAAGAAATACTAGCACAACCTGACA
GTGTATATACAGGTGGGACTGAAGTGATGGTACCTTCTTTCTGTAATCTGAAGAACCTGATT
CTATTACCAATCCATTAGTTTACCATCTGTTTCTCTGAACTATGGACAAACCTGTAGATT
TGCAACTAGAAAGGAAATTGATACAGATCTACAAGCCAAGGGGAAAGCAAGATAGTTTCAT
TTGGATTTGGAAGTAGCACAGGGCTCTCATTGTCAGACTTGGCTTCCAGTAATCTGGAGATT
TTGCTTTTGGTTCTAAAGATAAAAATTTCCAATGGGCAAACTAGGAGCAGCTGTGTTTGGAA
CACAGTCAGTCGGAACCCAGTCAGCCGGTAAAGTTGGTGAAGATGAAGATGGTAGTGAAG
AAGTAGTTCATAATGAAGATATCCATTTTGAACCAATAGTGTCACTACCAGAGGTAGAAGTAA
AATCTGGAGAAGAAGATGAAGAAATTTTGTTTAAAGAGAGAGCCAACTTTATAGATGGGATC
GGGATGTCAGTCAGTGAAGGAGCGCGGTGTTGGAGATATAAAGATTTCTTTGGCATAACAATGA
AGAATTATTACCGGATCTAATGAGAAGAGACCAGGTTTTTAAAGTGTGTGCAACCAACGTTA
TTACTAAAACAAATGGAATTAAGCCCTTAAATGTTTCAAATAATGCTTTAGTTTGGACTGCCT
CAGATTATGCTGATGGAGAAGCAAAGTAGAACAGCTTGCAGTGAGATTTAAAACCTAAAGAAG
TAGCTGATTGTTTCAAGAAAACATTTGAAGAATGTCAGCAGAATTTAATGAACTCCAGAAAG
GACATGTATCACTGGCAGCAGAAATATCAAAGGAGACCAATCTGTGGTGTTTTTGTGATGTTT
GTGCGACGGTGAACCTCTAGGCGGATAACTATGGAATTTATTTTCAAACATTTGTTCTCGGA
CTGCTGAGAACTTCAGAGCACTATGCACTGGAGAGAAAGGCTTTGGTTTCAAGAATTCATTT
TTCACAGAGTAATTCAGATTTTGTGTTGCAAGGAGAGATATCACCACCAATGATGGAACAG
GCGGACAGTCCATTTATGGAGACAAATTTGAAGATGAAAATTTTGTGTAACATACTGGTC
CTGGTTTACTATCCATGGCCCAATCAAGGCCAGAATACCAATAATCTCAATTTGTTTAAACAC
TGAAGAAAGCAGAACATTTGGACTTTAAGCATGTAGTATTGGGTTTGTAAAGATGGCATGG
ATACTGTGAAAAGATGAAATCATTGTTCTCCAAAGGGTCTGTTTGTGCAAGAAATAACTA
TCACAGAATGTGGACAGATATAAATCATTTGTTGTTTATAGAAAATTTTATCTGTATAAGCAG
TTGGATTGAAGCTTAGCTATTACAATTTGATAGTTATGTTTCACTTTTGAAGTGGACGTTT

	<p>CGATTTCACAAATGTAATAATGCAGCTTATAGCTGTGTCACTTTTAAATGTGTATAATTGAC CTTGATGGTGTGAAATAAAAGTTAAACACTGGTGTATTCAGGTGACTTGTGTTTATGTA CTCCTGACGTATTAATAATGAATAATACTAATCTTGTAAAGCAATAGACCTCAAACTATTG AAGGAATATGATATATGCAATTTAATTTAATTCCTTTAAGATATTGGACTTCTGTCATGG ATATACTTACCATTTGAATAAAGGACCACAACCTGGATAATTTAATTTTAGGTTTGAATAT ATTTGGTAATCTTAATACTATTGGTGTACTCATTATGCATAGAGACTCGTTTATGAATGGGTAG AGCCACAGAACGTATAGAGTTAACCAGTGTCTTCTCTAGAATCTTACACCTCCTGTGTG GTTACAAGTTAACTTTGTAAGTAGCGTACCTTCTTCTTAAATATCTAGCTTCTGTGCCC TTTCATAGATATTCGATTAAATTTTACATTTTAAACAAGTTGACTATTTCCCTTTAGGGGTTTT GTTTCAAACCTTTTCTGTCTCTGTCTCTACTACCTCAGAACTGCAGCTTGGTTCTGATGATA GAAATTGAATTTTCTTGTAGTTATTGTGATAAAGTATGAATATTTTAAAGAGTCTATACC ATGTTCTTTCGTTAAAGATTGCTTTATACAAGATTGTTGCAGTACCTTTTCTGGTAAATTT TGTAGCAGAAATAAATGACAATTCCTAAG</p>
	<p>ORF Start: ATG at 114 ORF Stop: TAA at 9786</p>
	<p>SEQ ID NO: 326 3224 aa MW at 358214.5kD</p>
NOV90a, CG94377-01 Protein Sequence	<p>MRRSKADVERYIASVQGSTPSRQKSMKGFYFALYYEAEKEYDLAKYICTYINVQERDPKAH RFLGLLYELEENTDKAVECYRRSVELNPTQKDLVLKIAELLCKNDVTDGRACYWLERAAKLP GSPAIYKLKEQLDCEGEDGWNKLFDLIQSELYVRPDDVHVNIRLVEVYRSTKRLKDAVAHCH EAERNIALRSSLEWNSCVVQTLKEYLESQCLESDKSDWRATNTDLLLAYANMLLTLSTRDV QESRELLQSFDALQSVKSLGGNDELSTFLEMKGHFYMHAGSLLLMKGQHSNNVQWRALSEL AALCYLIAFQVPRPKIKLIKGEAGQNLLEMMACDRLSQSGHMLNLSRGKQDFLKEIVETFFAN KSGQSALYDALFSSQSPKDTSLGSDDIGNIDVREPELEDLTRYDVGAIRAHNGSLQHLTWLG LQWNSLPALPGIRKWLKQLFHHLPHESTRLETNAPESICILDLEVLFLGVYVTSHLQLKEKN SHHSSYQPLCLPLPVCKQLCTERQKSWWDVAVCTLIHRKAVPGNVAKRLRLVQHEINTLRAQEK HGLQPALLVHWAECQKGTGSLNSFYDQREYIGRSVHYWKKVLPPLKIKKKNSIPEPIDPLF KHFSVVDIQASEIVEYEEDAHITFAILDVNGNIEDAVTAFESI KSVSVYWNALIFHRKAED IENDALSPREEQECKNYLRKTRDYLIKIIDSDSNLSVVKLPVPLESVKEMLSVMQLEDEY SEGGLYKNGSLRNADSEIKRSTPSPTRYSLSPSKSYKYSPKTPPRAWEDQNSLLKMICQQVE AIKKEMQELKLNSSNSASPHRWPTENYGPDSVPDGYQGSQTFHGAPLTVATTGPSVYYSQSPA YNSQYLLRPAANVTPTKGPVYGMNRLPPQHIYAYPQQMHTPPVQSSSACMFSSQMYGPPALR FESPATGILSPRGDDYFNYVQQTSTNPPLPEPGYFTKPPIAAHASRSAESKTI EFGKTNFVQ PMPGEGLRPSLPTQAHTTQPTPFKFSNFKSNDGDTFTSSPQVVTQPPPAAYSNSESLLGLLT SDKPLQGDGYSGAKPIPGGQTI GPRNTFNFGSKNVSGISFTENMGSSQKNSGFRSDDMFTF HGPGKSVFGTPTLETANKNHETDGGSAHGDDDDGPHFEPVVPDPKIEVKTGEDEEEFFCN RAKLFRFDVESKEWKERGIGNVKILRHKTSGKIRLLMRREQVLKICANHYSISPMKLTNPAGS DRSFVWHALDYADELPKPEQLAIRFKTPEEAALFKCKFEEAQSLKAPGTNVAMNSQAVRIV KEPTSHDNKDICKSDAGNLNFEFQVAKKEGSWWHCNSCSLKNASTAKKCVSCQNLNPSNKLIV GPPLAETVFTPKTSPENVQDRFALVTPKKEGHWDCSI CLVRNEPTVSRACQNTKSANKSGS SFVHQASFQGGDLPKPINSDFRSVFSTKEGQWDCSACLVRNEGSSTKCAACQNPQRKQSLPA TSIPTASFKFGTSETSKTLKSGFEDMFAKKEGQWDCSCLVRNEANATRCVACQNPDKPSPS TSVPAPASFQFGTSETSKAPKSGFEGMFTKKEGQWDCSVCLVRNEASATKCIACQNPQKQNT TSAVSTPASSETSKAPKSGFEGMFTKKEGQWDCSVCLVRNEASATKCIACQNPQKQNTTSAV STPASSETSKAPKSGFEGMFTKKEGQWDCSVCLVRNEASATKCIACQCPKQKQNTTASTPAS SEISKAPKSGFEGMFIRKQWDCSVCCVQNESSLKCVACDASKPTHKPIAEAPSFTLGESE KLHDSGSGSQVTGFKSNFSEKAKFGNTEQGFKFGHVDQENSPSFMFQGSNDSEITPHVSRSS IPVSADGFKFGISEPGNQEKSEKPLENGTGFAQDISGQKNGRGVIFGQTSSTFTFADLAKE TSGEGFQFGKDPNFKGFSGAGEKLFSSQYGMANKANTSGDFEKDDDAYKTEDSDDIHFEFV VQMPKVELVTGEEDEKVLVSQRVKLFRFDAEVSQWKERGLGNLILKNEVNGKRLMLMRREQ VLKVCANHWTITTMNLKPLSGSDRAWMWLASDFSDGDAKLEQLAAKFKTELAEEFKQKFECE QRLLLDIPLOTQPHKLVDTGRAAKLIQRAEEMKSGLKDFKFTFLNDQTKVTEENKSGTGAAG ASDTTIKPNPENTGPTLEWDNYDLREDALDDSVSSSVHASPLASSPVRKNLFRFGESTTGFN FSFKSALSPSKSPAKLNQSGTSGVTDEESDVTQEEERDQGYFEPVVPDLVEVSSGEENEQV VFSHRAKLYRYDKDVQWQKERGIGDIKILQNYDNKQVRIVMRRDQVLKCANHRIPTDMLQN MKGTERVWLWACDFADGERKVEHLAVRFKLQDVADSFKKIFDEAKTAQEKSDSEITPHVSRSS TPRESPCGKIAVAVLEETTRERTDVIQGGDVADATSEVEVSSTSETTPKAVVSPPKFVFGSES VKSIFSSEKSKPFAFGNSSATGSLFGFSFNAPLKSNNSETSSVAQSGSESKVEPKKCELSKNS DIEQSSDSKVKNLFAFPTEESSINYFTKTEKAKEKKKPEDSPSDDDLVIVELTPTAEQKA LATKLKLPPTFFCYKNRPDYVEEEDDEDFTAVKKNLGLYLDGSEKCRPLEENTADNEKE CIIVWEKKPTVEEKAKADTLKLPPTFFCGVCSDTDEDNGNGEDFQSELQKVQEAQKSQTTEEIT STTDSVYTGTEVMVPSFCKSEEPSITKSISSPSVSSETMDKPVDLSTRKEIDTDSQGES KIVSFGFGSSTGLSFADLASSNSGDFAFGSKDKNFQWANTGAAVFGTQSVGTQSGAGVGEDED GSDEEVVHNEDIHFEFIVSLPEVEVKSGEDEEELFKERAKLYRWDRDVSQWKERGVGDIKIL WHTMKNYRILMRDQVFKVCANHVIKTMLKPLNVSNNALVMTASDYADGEAKVEQLAVRF KTKEVADCFKKTFECCQNLMLKQGHVSLAAELSKETNPVVFDDVCADGEPLGRITMELFSN IVPRTAENFRALCTGEGFGFKNSIFHRVIPDFVCQGGDITKHDGTGGQSIYGDKFEDENFDV</p>

	KHTGPGLLSMANQGQNTNNSQFVITLKKAEHLDFKHVVFVKDGM DTVKKIESFGSPKGSVC RRITITECGQI		
	SEQ ID NO: 327	5332 bp	
NOV90b, CG94377-02 DNA Sequence	ACGCGTCTCGGGAGCCAGGTTGGCGGCGGATGAGGCGCAGCAAGGCCGATGTGGAGCGGTAC GTCGCCTCGGTGCTGGGTCTCACCCCGTCGCCTCGACAGAAGTCAATGAAAGGATTCTATTTT GCAAAGCTGTATTATGAAGCTAAAGAATATGATCTTGCTAAAAAGTACGTATGTACTTACCTT AGTGTGCAAGAGAGGGATCCCAGAGCTCACAGATTCTGGGTCTTCTTTATGAATTGGAAGAA AACACAGAGAAAGCCGTTGAATGTTACAGGCGTTCACTGGAATTAACCCACCACAAAAAGAT CTTGTTGAAGATTGCAGAATTGCTTTGTAATAATGATGTTACTGATGGAAGAGCAAAATAC TGGGTTGAAAGGCGAGCGAAACTTTTCCAGGAAGTCTGCAATTATATAAACTAAAGCATCTT CTAGATTGTGAAGGTGAAGATGGATGGAATAAACTTTTGTACTGGATTCACTCAGAACTTTAT GTAAGACCTGATGACGTCCATATGAACATCCGGCTAGTGGAGTTGATCGCTCAAATAAAGA TTGAAGGATGCTGTGGCCCGTCCCATGAGGCAGAGAGGAACATAGCTTTGCGTTCAAGTTTA GAGTGAATTCTGTGTGTGTACAGACCCTTAAGGAATATCTGGAGTCTTACAGTGTGGAG TCTGATAAAAGTGACTGGCGAGCAACCAATACAGACTTACTGCTGGCCTATGCTAATCTTATG CTTCTTACGCTTTCCACTAGAGATGTGCAGGAAAGTAGAGAATTACTGGAAGTTTGTAGT GCTCTTCAGTCTGCAAAATCTTCTTGGGTGGAATGATGAAGTGTGCACTTCTTTAGAA ATGAAAGGACATTTCTACATGCATGCTGTTCTCTGCTCTGAAGATGGGTGAGCATGGTAAT AATGTTCAATGGCAAGCTCTTCTGAGCTGGCTGCATTGTGCTATGTGCATAGCATTTCAGGT CCAAGACCAAAGATTAAATTAATAAAGGTGAAGCTGGACAAAATCTGCTGGAATGATGGCC TGTGACCGACTGAGCCAATCAGGCGCATATGTTGCTAAACTTAAGTCGTGGCAAGCAAGATTTT TTAAAGAGGTTGTTGAACTTTTGCCAAACAAAGCGGGCAGTCTGTGTTATATAATGCTCTG TTTTCTAGTCAGTCATCTAAGGATACATCTTTCTTGGTAGCGATGATATTGGAACATTGAT GTACAAGAACCAGAGCTTGAAGATTGGCTAGATACGATGTTGGTGCTATTCAAGCACATAAT GGTAGTCTTCAGCACCTTACTTGGCTTGGCTTACAGTGAATTCAATGCTGCTTTACCTGGA ATCCGAAAATGGCTAAACAGCTTTTCCATCATTGCCCCAGGAAACCTCAAGGCTTGAAACA AATGCACCTGAATCAATATGTATTTAGATCTTGAAGTATTTCTCCTTGGAGTAGTATATACC AGCCACTTACAATTAAGGAGAAATGTAATCTCACCACAGCTCCTATCAGCCGTTATGCCTG CCCCTTCTGTGTGTAACGGCTTTGTACAGAGAGACAAAATCTGGTGGGATGCGGTTTGT ACTCTGATTACAGAAAAGCAGTGAAGTCAAGCAGCAATGAGACTGTAGTTCAAGCATGAAATA AACACTCTAAGAGCCCAGGAAAACATGGCCTTCAACCTGCTCTGCTTGTACATTGGGCAAAA TGCCTTCAGAAAGGCAGGGGTCTTAATCTTCTTATGATCAACAAGAATACATAGGGAGAAGT GTTCAATTATTGGAAGAAAGTTTGCCATTGTTGAAGATAATAAAGAAGAACAGTATTCCTGAA CCTATTGATCCTCTGTTTAAACATTTTCATAGTGTAGACATTCAGGCATCAGAAATGTTGAG TATGAAGAAGATGCACACATACTTTTGCTATATTGGATGCAGTACATGGAATATAGAAGAT GCTGTGACTGCTTTGAATCTATAAAAGTGTGTTTCTTATTGGAATCTGCACTGATTTT CACAGGAAAGCAGAAAGACATTGAAATGATGCCGTTTTCTGAAGAACAGAAGAAATGCAAA AATTATCTGAGAAAGACCAGGACTACCTAATAAAGATTATAGATGACAGTGATTCAAATCTT TCAGTGGTCAAGAAAGTAAGTGTGCCCTGGAGTCTGTAAGAGAGATGCTTAAGTCAGTCAAG CAGGAATCGAAGACTATAGTGAAGGAGGTCTCTCTATAAAAATGGTTCTTTGCGAAATGCA GATTACAGAAATAAACATTTCTACACCATCTCCTACCAATATTCACTATACCAAGTAAAGT TACAAGTATTTCTCCAAAACACCACCTCGATGGGCAGAGATCAGAAATCTTTACGGAAAATG ATTTTGCCAAAGATAAGGCCATTAAAGAAAGAAATGCAGGAGTTGAAACTAAATAGCAGTAAG TCAGCATCCCGTCATCGTTGGCCACAGAGAATTATGGACCAGACTCGGTGCCTGATGGATAT CAGGGGTACAGACATTTATGGGGCTCCACTAACAGTTGCAACTACTGGCCCTCAGTATAT TATAGTCAGTCACCAGCATATAATTCCAGTATCTTCTCAGACCAGCAGCTAATGTTACTCCC ACAAAGGGTTCTTCTAATACAGAATTTAAGTCAACCAAGAAGGATTTTCCATCCCTGTGCT GCTGATGGATTAAATTTGGCATTTCGGAACCAGGAAATCAAGAAAAGAAAAGTGAAGAAGCCT CTTGAAAATGATACTGGCTTCCAGGCTCAGGATATTAGTGGCCAGAAGATGGCCGTGGTGTG ATTTTGGCCAAACAGTAGCACTTTTACATTGTCAGATGTTGCAAAATCAACTTCAGGAGAA GGATTTCAAGTTTGGCAAAAAGACCCCAATTTCAAGGGATTTTCAGGTGCTGGAGAAAATTA TTCTCATCACAATGCGGTAAATGGCCAATAAAGCAAACTTCCGGTACTTTGAGAAAGAT GATGATGCCTGTAAGACTGAGGACAGCGATGACATCCATTTGAACAGTAGTTCAAAATGCCT GAAAAGTAGAAGTTGTAACAGGAGAAGAAGTGAAAAGTTCTGTATTACAGGGGTAAAA CTATTTAGATTTGATGCTGAGATAAGTCAGTGGAAGAAAGGGCTTGGGGAACCTAAAAATT CTCAAAAATGAGGTCAATGGCAAACCAAGATGCTGATGCGAAGAGACCAAGTACTAAAAGTG TGTGCTAATCATTGGATAACAACATACTAATGAACCTGAAGCCCTCTCTGGATCAGATAGACA TGGATGTGGTTAGCCAGTGATTTCTGATGGTGATGCCAACTAGAGCGGTTGGCAGCACAA TTTAAACACCCAGAGCTGGCTGAAGAATTCAAGCAGAAATTTGAGGAATGCCAGCGGCTCTG TTAGACATACCACCTTCAAACTCCCCATAAACTTGTAGATACTGGCAGAGCTGCCAAGCTAATA CAGAGAGCTGAAGAATGAAGATGGACTGAAAGATTTCAAAACGTTTTTGACAATGATCAA ACAAAAGTCACTGAGGAAGAAAATAAGGGTTCAAGGTACAGGTGCAGCCGGTGCCTCAGACACA ACAATAAAACCAATCTGAAAACACTGGGCCACATTAGAATGGGATAACTATGATTTAAGG GAAGATGCTTTGGATGATAATGTTAGTAGTAGCTCAGTACATGATTTCCGTTGGCAAGTAGC CTGTGAGAAAAATATTTCCGCTTTGATGAGTCAACAACAGGATTTAACTTCAGTTTAAA		

	TCTGCTTTGAGTCTATCTAAGTCTCCTGCCAAGTTGAATCAGAGTGGGACTTCAGTTGGCACT GATGAAGAATCTGATGTTACTCAAGAAGAAGAGAGAGATGGACAGTACTTTGAACCTGTTGTT CCTTTACCTGATCTAGTTGAAGTATCCAGTGGTGAGGAAAATGAACAAGTTGTTTTAGTCAG ATGGCAGAACCTTACAGATATGATAAAGATGTTGGTCAATGGAAAAGAAAGGGGCATTGGTGAT ATAAAGATTTTACAGAATTATGATAATAAGCAAGTTCGTATAGTGATGAGAAGGGACCAAGTA TTAAACTTTGTGCCAATCACAGAATAACTCCAGACATGAGTTTGCAAAATATGAAAGGGACA GAAAGAGTATGGGTGTGGACTGCATGTGATTTTGAGATGGAGAAAAGAAAAGTAGAGCATTTA GCTGTTTCGTTTTAACTACAGGATGTTGCAGACTCATTAAAGAAAATTTTGATGAAGCAAAA ACAGCCCAGGAAAAAGATTCTTTGATAACACCTCATGTTTCTCGGTCAAGCACTCCCAGAGAG TCACCATGTGGCAAAATTGCTGTAGCTGTATTAGAAGAAACCACAAGAGAGAGGACAGATGTT ATTCAGGGTGATGATGTAGCAGATGCAGCTTCAGAAGTTGAAGTGTCTAGCACATCTGAAACA ACAACAAAAGCAGTGGTTTCTCCTCCAAAGTTTGATTTGGTTTCAGAGTCTGTTAAAGAATT TTTAGTAGTGAAAAATCAAACCCATTTGCATTGGCAACAGTTCTGCCACTGGGTCTTTGTTT GGATTTAGTTTTAATGCACCTTTGAAAAGTAACGATAGTGAAGTCTAGTACGCCAGAGT GGATCTGAAAGCAAAGTGAACCTAAAAATGTGAAGTGTCAAAGAACTCTGATATCGAACAC TCTTCAGATAGCAAAAGTCAAAAATCTCTCTGCTTCTTTCCAATGGAAGAATCTTCAATCAAG TACACATTTAAACACCAGAAAAGGAGCCTCCATTATGGCATGCTGAATTTACCAAGAAGAA TTGGTTTCAAGAGCTCAGTTCCACCACAAAAGTGCAGATCAGTTAAACGGCCTGCTTCGGGAA ACAGAGGCAACACAGTGCAGTCTTATGGAGCAAAATTAAGCTTCTCAAAGTGAAATAAGAAGA TTGGAAGGAATCAAGAGGAGTCTGCAGCTAACGTGGAACACTTGAAGAAGCTCTTGTGTCAG TTCATTTTCTTGAAGCCAGGTAGTGAGAGAGAGAGCCTTCTTCTGTTATAAATACGATGTTG CAGCTCAGCCCTGAAGAAAAGGAAAACCTGCTGCGGTGCTCAAGGTCTTCAACAAACCTCC ATACCCAAGAAAAAATAGAAAGCACCATTGTTCTACTATGG		
	ORF Start: at 1		ORF Stop: TAG at 5308
	SEQ ID NO: 328	1769 aa	MW at 198802.2kD
NOV90b, CG94377-02 Protein Sequence	TRLGSSQVGGAMRRSKADVERYVASVLGLTPSPRQKSMKGFYFAKLYEAEKDYLAKEYVCTYL SVQERDPRAHRFLGLLYLEEENTEKAVECYRRSLELNPPQKDLVLKIAELLCKNDVTDGRAKY WVERAAKLFPGSPAIIYKLKHLDDCEGEDGWNKLFDWIQSELYVRPDDVHMNIRLVELYRSNKR LKDAVARCHEAERNIALRSSLEWNSCVVQLKEYLESQLESKSDWRATNTDLLAYANLM LLTLSTRDVQESRELLESFDSALQSAKSSLGNDLSATFLEMKGHFYMHAGSLLLKMGQHG NVQWQALSELALCYVIAFOVPRPKIKLIKGEAGQNLLEMMACDRLSQSGHMLNLSRGKQDF LKEVVETFANKSGQSVLYNALFSSQSSKDTSLGSDDIGNIDVQEPEDLARYDVYIAQAHN GSLQHLTWLGLQWNSLPALPGIRKWLKQLFHHLPQETSRLTNAPESICILDLEVFLLGVVYT SHLQLKEKCNSSHSSYQPLCLPLPVCRLCTERQKSWWDVCTLIHRKAVNSAELRLVVQHEI NTLRAQEKHGLQFALLVHWAKCLQKGRGLNSSYDQOEYIGRSVHYWKKVLPPLKIIKKNSIPE PIDPLFKHFSVDIQASEIVEYEEDAHITFAILDVHGNIEDAVTAFESI KSVSVYWNALIF HRKAEDIENDAVFPPEEQECKNYLRKTRDYLIKI IDSDSNLSVVKVSVPLESVKEMLKSV QELEDYSEGGPLYKNGSLRNADSEIKHSTPSPTKYSLSPSKSYKYSKPTPPRWAEDQNSLRKM ICQEVKAIKKEMQELKLNSSKSASRHRWPTENYGPDSVPDGYQGSQTFHGAPLTVATTGPSVY YSQSPAYNSQYLLRPAANVTPTKGSNTEFKSTKEGFSIPVSADGFKFGISEPGNQEKSEK LENDTGFAQADISGQKNGRGVIFGQTSSTFTFADVAKSTSGEGFQFGKDPNFKGFSGAGEKL FSSQCGKMANKANTSGDFEKDDACKTEDSDDIHFEPPVQMPEKVELVTGEEGKVLVSQGVK LFRFDAEISQWKERGLNLKILKNEVNGKPRMLMRRDQVLKVCANHWITTTMNLKPLSGSDRA WMLASDFSDGAKLERLAAQFKTPELAEFEKQKFEECQRLLLDIPLQTPHKLVDTGRAAKLI QRAEEMKSGLKDFKTFLTNDQTKVTEENKSGTGAAGASDTTIKPNPENTGPTLEWDNYDLR EDALDDNVSSSSVHDSPLASSPVRKNIFRDESTTGFNFSFKSALSLSKSPAKLNQSGTSVGT DEESDVTQEEERDQYFEPVVPDLVEVSSGEENEQVVFSHMAELYRYDKDVGQWKERGIGD IKILQNYDNKQVRIVMRRDQVLKLCANHRITPMSLQNMKGTERVWVWTAACDFADGERKVEHL AVRFLQDVADSFKKIFDEAKTAQEKDSLITPHVSRSTPRESPCGKIAVAVLEETTRERTDV IQGDDVADAASEVEVSSTSETTTKAVVSPPKFVFGSESVKRIFSEKSNPFAFGNSSATGSLF GFSFNAPLKSNDSETSSVAQSGSESKVEPKKCELSKNSDIEQSSDSKVKNLSASFPMEESSIN YTFKTPEKEPPLWHAFTKEELVQKLSSTTKSADQLNGLLRETEATSAVLMEQIKLLKSEIRR LERNQESAAANVEHLKNVLLQFIFLKPGRSERESLLPVINTMLQLSPEEKGLAQAQGLQOTS IPKKK		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 90B.

Table 90B. Comparison of NOV90a against NOV90b.

Protein Sequence	NOV90a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV90b	1..900 11..906	795/901 (88%) 823/901 (91%)

Further analysis of the NOV90a protein yielded the following properties shown in Table 90C.

Table 90C. Protein Sequence Properties NOV90a

PSort analysis:	0.6000 probability located in endoplasmic reticulum (membrane); 0.3000 probability located in microbody (peroxisome); 0.2525 probability located in mitochondrial inner membrane; 0.1000 probability located in nucleus
SignalP analysis:	No Known Signal Sequence Predicted

5

A search of the NOV90a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 90D.

Table 90D. Geneseq Results for NOV90a

Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV90a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAW54235	Human Nup358 protein - Homo sapiens, 3224 aa. [WO9809170-A2, 05-MAR-1998]	1..3224 1..3224	3224/3224 (100%) 3224/3224 (100%)	0.0
AAM03867	Peptide #2549 encoded by probe for measuring breast gene expression - Homo sapiens, 164 aa. [WO200157270-A2, 09-AUG-2001]	1885..2048 1..164	157/164 (95%) 159/164 (96%)	2e-85
AAM28631	Peptide #2668 encoded by probe for measuring placental gene expression - Homo sapiens, 164 aa. [WO200157272-A2, 09-AUG-2001]	1885..2048 1..164	157/164 (95%) 159/164 (96%)	2e-85
AAM16137	Peptide #2571 encoded by probe for measuring cervical gene expression - Homo sapiens, 164 aa. [WO200157278-A2, 09-AUG-2001]	1885..2048 1..164	157/164 (95%) 159/164 (96%)	2e-85
AAM68322				2e-85

probe encoded protein SEQ ID NO: 28628 - Homo sapiens, 164 aa. [WO200157276-A2, 09-AUG-2001]	1..164	159/164 (96%)	
--	--------	---------------	--

In a BLAST search of public sequence databases, the NOV90a protein was found to have homology to the proteins shown in the BLASTP data in Table 90E.

Table 90E. Public BLASTP Results for NOV90a				
Protein Accession Number	Protein/Organism/Length	NOV90a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
P49792	Ran-binding protein 2 (RanBP2) (Nuclear pore complex protein Nup358) (Nucleoporin Nup358) (358 kDa nucleoporin) (P270) - Homo sapiens (Human), 3224 aa.	1..3224 1..3224	3224/3224 (100%) 3224/3224 (100%)	0.0
S58884	Ran-binding protein 2 - human, 3224 aa.	1..3224 1..3224	3222/3224 (99%) 3223/3224 (99%)	0.0
Q9ERU9	Ran-binding protein 2 - Mus musculus (Mouse), 3053 aa.	1..1656 1..1670	1406/1687 (83%) 1505/1687 (88%)	0.0
P48820	Ran-binding protein 2 (RanBP2) (Nuclear pore complex protein Nup358) (Nucleoporin Nup358) (358 kDa nucleoporin) (P270) - Bos taurus (Bovine), 1085 aa (fragment).	2136..3224 1..1085	944/1090 (86%) 998/1090 (90%)	0.0
Q99666	Sperm membrane protein BS-63 - Homo sapiens (Human), 1765 aa.	1..900 1..900	853/901 (94%) 875/901 (96%)	0.0

5

PFam analysis predicts that the NOV90a protein contains the domains shown in the Table 90F.

Table 90F. Domain Analysis of NOV90a			
Pfam Domain	NOV90a Match Region	Identities/ Similarities for the Matched Region	Expect Value
TPR	60..93	11/34 (32%) 27/34 (79%)	7.1e-07
Ran_BP1	1183..1304	87/127 (69%) 120/127 (94%)	3.6e-90
zf-RanBP	1351..1381	15/32 (47%) 27/32 (84%)	1.5e-08

zf-RanBP	1415..1444	15/32 (47%) 26/32 (81%)	1.2e-10
zf-C3HC4	1485..1502	7/26 (27%) 16/26 (62%)	0.76
zf-RanBP	1479..1508	18/32 (56%) 28/32 (88%)	3.3e-12
zf-RanBP	1543..1572	18/32 (56%) 28/32 (88%)	8.9e-12
zf-RanBP	1606..1635	17/32 (53%) 29/32 (91%)	1e-12
zf-RanBP	1665..1694	17/32 (53%) 29/32 (91%)	1e-12
zf-RanBP	1724..1753	17/32 (53%) 28/32 (88%)	1.1e-10
zf-RanBP	1781..1810	19/32 (59%) 29/32 (91%)	1.7e-12
Ran_BP1	2024..2145	86/127 (68%) 120/127 (94%)	9.7e-88
Ran_BP1	2321..2442	77/127 (61%) 121/127 (95%)	1.1e-85
Ran_BP1	2922..3043	84/127 (66%) 122/127 (96%)	6.8e-92
pro_isomerase	3065..3224	100/179 (56%) 140/179 (78%)	1.7e-90

Example 91.

The NOV91 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 91A.

5

Table 91A. NOV91 Sequence Analysis			
	SEQ ID NO: 329	1908 bp	
NOV91a. CG97090-01 DNA Sequence	ACAGGTGACTTTCCACAGGAACCTTCTGCAATGTCCCATCAACCTCTCAGCTGCTGGAATTG CCCTTATCCTCCACCTGGATCTCCCAAACCTGGACACATTTACCCGGAGGAGCTGCTGCAG CAGATGAAAGAGCTCCTGACCGAGAACCACAGCTGAAAGAAGCCATGAAGCTAAATAATCAA GCCATGAAAGGAGATTGTAGGAGCTTCGGCCTGGACAGAGAAACAGAAAGGAAGAACGCCAG TTTTTTGAGATACAGAGCAAGAAGCAAAAGAGCGTCTAATGGCCTTGAGTCATGAGAATGAG AAATTGAAGGAAGAGCTTGGAAACTAAAAGGGAAATCAGAAAGGTCATCTGAGGACCCCACT GATGACTCCAGGCTTCCCAGGCGCAAGCGGAGCAGGAAAAGGACCAGCTCAGGACCCAGGTG GTGAGGCTACAAGCAGAGAAGGCAGACCTGTTGGGCATCGTGTCTGAACTGCAGCTCAAGCTG AACTCCAGCGGCTCCTCAGAAGATTCCTTTGTTGAAATTAGGATGGCTGAAGGAGAAGCAGAA GGGTCAGTAAAAGAAATCAAGCATAGTCTGGGCCACGAGAACAGTCTCCACTGGCAGCAGC AGATCTGCAGATGGGGCCAAGAATTACTTCGAACATGAGGAGTTAACTGTGAGCCAGCTCCTG CTGTGCCTAAGGGAAGGGAATCAGAAGGTGGAGAGACTTGAAGTTGCACTCAAGGAGGCCAAA GAAAGAGTTTCAGATTTTGAAAGAAAACAAGTAATCGTTCTGAGATTGAAACCCAGACAGAG GGGAGCACAGAGAAAGAGAATGATGAAGAGAAAGGCCCGAGACTGTTGGAAGCGAAGTGAA		

	GCACTGAACCTCCAGGTGACATCTCTGTTTAAGGAGCTTCAAGAGGCTCATACAAAACCTCAGC GAAGCTGAGCTAATGAAGAAGAGACTTCAAGAAAAGTGTGAGGCCCTTGAAAGGAAAAATTCT GCAATTCATCAGAGTTGAATGAAAAGCAAGAGCTTGTATTACTAACAAAAAGTTAGAGCTA CAAGTGGAAGCATGCTATCAGAAATCAAAATGGAACAGGCTAAAAACAGAGGATGAAAAGTCC AAATTAACGTGTACAGATGACACACAACAAGCTTCTTCAAGACATAATAATGCATTGAAA ACAATTGAGGAATAACAAGAAAAGAGTCAGAAAAGTGGACAGGGCAGTGTGAAGGAACTG AGTGAAAAACTGGAAGTGGCAGAGAAGGCTCTGGCTTCCAAACAGCTGCAATGGATGAAATG AAGCAAACCATTGCCAAGCAGGAAGAGGACCTGGAACCATGACCATCCTCAGGGCTCAGATG GAAGTTTACTGTTCTGATTTTCATGCTGAAAGAGCAGCGAGAGAGAAAATTATGAGGAAAAG GAGCAACTGGCATTGCAGCTGGCAGTTCTGCTGAAAGAGAATGATGCTTTCGAAGACGGAGGC AGGCAGTCCTTGATGGAGATGCAGAGTCGTCATGGGGCGAGAACAAGTGACTCTGACCAGCAG GCTTACCTTGTTCAAAGAGGAGCTGAGGACAGGGACTGGCGGCAACAGCGGAATATTCCGATT CATTCTGCCCCAAGTGTGGAGAGGTTCTGCCTGACATAGACACGTTACAGATTACAGTGATG GATTGCATCATTTAAGTGTGATGTATCACCTCCCCAAAACCTGTTGGTAAATGTCAGATTTTT TCCTCCAAGAGTTGTGCTTTTGTGTTATTGTTTTCACTCAAATATTTGCTCATTATTCTT GTTTTAAAAGAAAGAAAACAGGCCGGGCACAGTGGCTCATGCCTGTAATCCAGCACTTTGGG AGATCCAGGTGGGAGGAT		
	ORF Start: ATG at 31		ORF Stop: TAA at 1714
	SEQ ID NO: 330	561 aa	MW at 64267.6kD
NOV91a, CG97090-01 Protein Sequence	MSHQPLSCWNSPLSSHLDLNPLDTFTPEELLQOMKELLTENHQLKEAMKLNQAMKGRFEELS AWTEKQKEERQFFEIQSKEAKERLMALSHENELKEELGKLKGKRSERSEDPTDDSRLPRAEA EQEKDQLRTQVRLQAEKADLLGIVSELQKLNSGSSSEDSFVEIRMAEGEAEGSVKEIKHSP GPTRTVSTGTSRSADGAKNYFEHVELTVSQQLLCLREGNQKVERLEVALKEAKERVDFEKK SNRSEIETQTEGSTKENDEEKGPETVGSEVEALNLQVTSFLKELQEAHTKLSEALMKRRLQ EKQALERKNSAIPSELNEKQELVYTNKKLELQVESMLSEIKMEQAKTEDEKSKLTVLQMTN KLLQEHNNALKTIEELTRKESEKVDRAVLKELSEKLELAELALASKQLQMDQEMKQTIKQEE LETMTILRAQMEVYCSDFHAERAAREKIHKEEQALQLAVLLKENDAFEDGGRQSLMEMQSR HGARTSDSDQAYLVQRGAEDRDWRQQRNIPHSCKPCKGEVLPDIDTLQIHVMDCII		
	SEQ ID NO: 331	1858 bp	
NOV91b, CG97090-04 DNA Sequence	ATCCTCCACCTGGATCTCCCAACCTGGACACGTTTACCCCGAGGAGCTGCTGCAGCAGAT GAAAGAGCTCCTGACCGAGAACCACAGCTGAAAGAAGCCATGAAGCTAAATAATCAAGCCAT GAAAGGGAGATTGAGGAGCTTTTCGGCCTGGACAGAGAAACAGAAAGAACGCCAGTTT TGAGATACAGAGCAAAGAAGCAAGAGCGTCTAATGGCCTTGAGTCATGAGAATGAGAAAT GAAGGAAGAGCTTGGAAAACCTAAAGGGGAAATCAGAAAGTTCATCTGAGGACCCACTGATGA CTCCAGGCTTCCAGGGCCGAAGCGGAGCAGGAAAAGGACCAGCTCAGGACCCAGGTGGTGAG GCTACAAGCAGAGAAGGCAGACCTGTTGGGCATCGTGTCTGAAGTGCAGCTCAAGCTGAAGTC CAGCGGCTCCTCAGAAGATTCCTTGTGTAATAGGATGGCTGAAGGAGAAGCAGAAGGGT AGTAAAAGAAATCAAGCATAGTCTGGGCCACGAGAACAGTCTCACTGGCAGCGCATTGTC TAAATATAGGAGCAGATCTGCAGATGGGGCCAAGAATTACTTCGAACATGAGGAGTTAACTGT GAGCCAGCTCCTGCTGTGCTTAAGGGAAGGGAATCAGAAGGTGGAGAGACTTGAAGTTGCACT CAAGGAGGCCAAAGAAAGAGTTTCAAGTTTGAAGAAAGAAACAAGTAATCGTTCTGAGATTGA AACCAGACAGAGGGGAGCACAGAGAAGAGAAATGATGAAGAGAAAGGCCCGGAGACTGTTGG AAGCGAAGTGAAGCACTGAACCTCCAGGTGACATCTCTGTTTAAAGGAGCTTCAAGAGGCTCA TACAAAACCTCAGCGAAGCTGAGCTAATGAAGAAGAGACTTCAAGAAAAGTGTGAGGCCCTTGA AAGGAAAAATCTGCAATTCATCAGAGTTGAATGAAAAGCAAGAGCTTGTATTACTAACAA AAAGTTAGAGCTACAAGTGAAGCATGCTATCAGAAATCAAAATGGAACAGGCTAAAACAGA GGATGAAAAGTCCAAATTAAGTGTGCTACAGATGACACACAACAAGCTTCTTCAAGAATATAA TAATGCATTGAAAACAATTGAGGAATAACAAGAAAAGAGTCAGAAAAAGTGGACAGGCAGT GCTGAAGGAAGTGAAGTGAAGTGAAGTGGCAGAGAAGGCTCTGGCTTCCAAACAGCTGCA AATGGATGAAATGAAGCAACCATTTGCCAAGCAGGAAGAGGACCTGGAAACCATGACCATCCT CAGGGCTCAGATGGAAGTTTACTGTTCTGATTTTCATGCTGAAAGAGCAGCGAGAGAGAAAA TCATGAGGAAAAGAGAGCACTGCATTGCAGCTGGCAGTCTGCTGAAAGAGAGATGATGCTT CGAAGACGGAGGCAGGCAGTCTTGATGGAGATGCAGAGTCGTCATGGGGCGAGAACAAGTGA CTCTGACCAGCAGGCTTACCTTGTTCAAAGAGGAGCTGAGGACAGGGACTGGCGGAACAGCG GAATATTCCGATTCACTCCTGCCCAAGTGTGGAGAGGTTCTGCTGACATAGACACGTTACA GATTCAGTGATGATTGCATCATTTAAGTGTAAATGTATCACCTCCCCAAAACCTGTTGGTAA ATGTCAGATTTTTCCTCCAAGAGTTGTGCTTTGTGTTATTGTTTCACTCAAATATTTTG CCTCATTATTCTTGTTTTAAAGAAAGAAAACAGGCCGGGCACAGTGGCTCATGCCTGTAATC CCAGCACTTTGGGAGATCCAGGTGGGAGGAT		
	ORF Start: ATG at 62		ORF Stop: TAA at 1664
	SEQ ID NO: 332	534 aa	MW at 61225.3kD
NOV91b,	MKELLTENHQLKEAMKLNQAMKGRFEELSAWTEKQKEERQFFEIQSKEAKERLMALSHENEL LKEELGKLKGKRSERSEDPTDDSRLPRAEAQEKDQLRTQVRLQAEKADLLGIVSELQKLNL		

CG97090-04 Protein Sequence	SSGSSSEDSFVEIRMAEGEAEAGSVKEIKHSPGPTRTVSTGTALSKYRSRSADGAKNYFEHEELT VSQLLLLCLREGNQKVERLEVALKEAKERVSDFEKKTNSRSEIETQTEGSTENDEEKGPETV GSEVEALNLQVTSFLKELQEAHTKLSEAEMLKKRLQEKQALERKNSAIPSELNEKQELVYTN KKLELQVESMLSEIKMEQAKTEDEKSKLTVLQMTNKLQEHNNALKTIEELTRKESEKVDRA VLKELSEKLELAEKALASKQLQMDQKQTIKQEEDELEMTILRAQMEVYCSDFHAERAAREK IHEEKEQLALQLAVLLKENDAFEDGGRQSLMEMQSRHGARTSDSDQAYLVQRGAEDRDWRQQ RNIPHSCPKCGEVLDPDIDTLQIHVMDCII		
	SEQ ID NO: 333	1857 bp	
NOV91c, CG97090-03 DNA Sequence	TGCTGGAATTCGCCCTTATCCTCCACCTGGATCTCCCAAACCTGGACACATTTACCCCGGAG GAGCTGCTGCAGCAGATGAAAGAGCTCCTGACCGAGAACCACAGCTGAAAGAAGCCATGAAG CTAAATAATCAAGCCATGAAAGGAGATTGAGGAGCTTTGGCCTGGACAGAGAACAGAAAG GAAGAAGCGCAGTTTGTGAGATACAGAGCAAAGAAGCAAAGAGCGCTTAATGGCCTTGAGT CATGAGAATGAGAAATTGAAGGAAGAGCTTGGAAAATAAAAGGGAAATCAGAAAGGTCTCT GAGGACCCCACTGATGACTCCAGGCTTCCAGGGCCGAAGCGGAGCAGGAAAAGGACAGCTC AGGACCCAGGTGGTGAGGCTACAAGCAGAGAAGGCAGACCTGTTGGGCATCGTGTCTGAAGT CAGCTCAAGCTGAAGTCCAGCGGCTCCTCAGAAGATTCTTTGTTGAAATTAGGATGGCTGAA GGAGAAGCAGAAGGGTCAGTAAAGAAATCAAGCATAGTCTGGGCCACGAGAACAGTCTCC ACTGGCACGAGCAGATCTGCAGATGGGGCCAAGATTACTTGAACATGAGGAGTTAACTGTG AGCCAGCTCCTGCTGTGCCTAAGGGAAGGGAATCAGAAGGTGGAGAGACTTGAAGTTGCACCT AAGGAGGCCAAAGAAAGAGTTTCAAGTTTGAAGAAAGAAACAAGTAATCGTTCTGAGATTGAA ACCCAGACAGAGGGGAGCACAGAGAAAGAGAATGATGAAGAGAAAGGCCCGGAGACTGTTGGA AGCGAAGTGGAAGCACTGAACCTCCAGGTGACATCTCTGTTAAGGAGCTTCAAGAGGCTCAT ACAAAATCAGCGAAGCTGAGCTAATGAAGAAGAGACTTCAAGAAAAGTGTGAGGCCCTTGAA AGGAAAATTTCTGAATTCATCAGAGTTGAATGAAAAGCAAGAGCTTGTTTATACTAACAAA AAGTTAGAGCTACAAGTGGAAAGCATGCTATCAGAAATCAAAATGGAACAGACTAAACAGAG GATGAAAAGTCCAAATTAAGTGTCTACAGATGACACACAACAGCTTCTTCAAGAACATAAT AATGCATTGAAAACAAATGAGGAAGTAAAGAAAGAGTCAAGAAAAGTGGACAGGCGAGTG CTGAAGGAAGTGAAGTGAAGTGAAGTGGCAGAGAGGCTCTGGCTTCCAAACAGCTGCAA ATGGATGAAATGAAGCAACCATTCGCAAGCAGGAAGAGGACCTGGAAACCATGACCATCTC AGGGCTCAGATGGAAGTTACTGTTCTGATTTTCATGCTGAAAGAGCAGCGAGAGAGAAAATT CATGAGGAAAAGGAGCAACTGGCATTGCAGCTGGCAGTTCTGCTGAAAGAGAATGATGCTTTC GAAGACGGAGGAGGAGTCTTATGATGAGATGCAGAGTCGTATGGGGCGAGAACAGTGAC TCTGACCAGCAGGCTTACCTTGTTCAAAGAGGAGCTGAGGACAGGACTGGCGGCAACAGCGG AATATTCGATTCATTCTGCCCCAAGTGTGGAGAGGTTCTGCCTGACATGACACGTTACAG ATTACGCTGATGGAATTGCATCATTTAAGTGTGATGATCACCTCCCCAAACTGTTGGTAAA TGTCAGATTTTTCTCCTCAAGAGTTGTGCTTTTGTGTTATTGTTTCACTCAAATATTTTGC CTCATTATTCTTGTGTTTAAAGAAAGAAAACAGGCCGGGCACAGTGGCTCATGCCTGTAATCC CAGCACTTGGGAGATCCAGTGGGAGGAT		
	ORF Start: ATG at 79		ORF Stop: TAA at 1663
	SEQ ID NO: 334	528 aa	MW at 60506.5kD
NOV91c, CG97090-03 Protein Sequence	MKELLTENHQLKEAMKLNQAMKGRFEELSAWTEKQKEERQFFEIQSKEAKERLMALSHENK LKEELGKLKGSERSSDDPTDSSLRAEAEQEKDQLRTQVVRLOAEKADLLGIVSELQKLNL SSGSSSEDSFVEIRMAEGEAEAGSVKEIKHSPGPTRTVSTGTALSKYRSRSADGAKNYFEHEELTVS CLREGNQKVERLEVALKEAKERVSDFEKKTNSRSEIETQTEGSTENDEEKGPETVVGSEVEA LNLQVTSFLKELQEAHTKLSEAEMLKKRLQEKQALERKNSAIPSELNEKQELVYTNKKLELQ VESMLSEIKMEQAKTEDEKSKLTVLQMTNKLQEHNNALKTIEELTRKESEKVDRAVLKELS EKLELAEKALASKQLQMDQKQTIKQEEDELEMTILRAQMEVYCSDFHAERAAREKIHKEE QLALQLAVLLKENDAFEDGGRQSLMEMQSRHGARTSDSDQAYLVQRGAEDRDWRQQRNIPH SCPKCGEVLDPDIDTLQIHVMDCII		
	SEQ ID NO: 335	1908 bp	
NOV91d, CG97090-02 DNA Sequence	ACAGGTGACTTTCCACAGGAACCTCTGCAATGTCCCATCAACCTCTCAGATCCTCCACCTG GATCTCCCAAACCTGGACACGTTTACCCCGGAGGAGCTGCTGCAGCAGATGAAAGAGCTCCTG ACCGAGAACCACAGCTGAAAGAAGCCATGAAGCTAAATAATCAAGCCATGAAAGGGAGATT GAGGAGCTTTTCGGCTGGACAGAGAAACAGAAAGGAAGAACGCCAGTTTGTGAGATACAGAGC AAAGAAGCAAAGAGCGCTTAATGGCCTTGAGTCAATGAGAATGAGAAATTGAAGGAAGAGCTT GGAAAATAAAAGGGAAATCAGAAAGGTCTCTGAGGACCCCACTGATGACTCCAGGCTTCCC AGGGCCGAAGCGGAGCAGGAAAAGGACAGCTCAGGACCCAGGTGGTGAGGCTACAAGCAGAG AAGGCAGACCTGTTGGGCATCGTGTCTGAAGTGCAGCTCAAGCTGAAGTCCAGCGGCTCCTCA GAAGATTCTTTGTTGAAATTAGGATGGCTGAAGGAGAAGCAGAAAGGTCAGTAAAGAAATC AAGCATAGTCTGGGCCACAGAAACAGTCTCCACTGGCACGGCATTGTCTAAATATAGGAGC AGATCTGCAGATGGGGCCAAGATTACTTGAACATGAGGAGTTAACTGTGAGCCAGCTCCTG CTGTGCCTAAGGGAAGGGAATCAGAAGGTGGAGAGACTTGAAGTTGCACTCAAGGAGGCCAAA GAAAGAGTTTCAGATTTTGAAGAAAGAAACAAGTAATCGTTCTGAGATTGAACCCAGACAGAG		

	GGGAGCACAGAGAAAGAGAATGATGAAGAGAAAGGCCCGGAGACTGTTGGAAGCGAAGTGGAA GCACTGAACCTCCAGGTGACATCTCTGTTTAAGGAGCTTCAAGAGGCTCATACAAAACCTCAGC GAAGCTGAGCTAATGAAGAAGAGACTTCAAGAAAAGTGTGAGGCCCTTGAAGGAAAAATTCT GCAATTCCATCAGAGTTGAATGAAAAGCAAGAGCTTGTTTATACTAACAAAAAGTTAGAGCTA CAAGTGGAAAGCATGCTATCAGAAATCAAAATGGAACAGGCTAAAAACAGAGGATGAAAAGTCC AAATTAAGTGTGCTACAGATGACACACAACAGCTTCTTCAAGAACATAAATATGCATTGAAA ACAATTGAGGAACATAACAAGAAAAGAGTCAGAAAAAGTGGACAGGGCAGTGTGAAAGAACTG AGTAAAAAAGTGAAGTGGCAGAGAAGGCTCTGGCTTCCAAACAGCTGCAATGGATGAAATG AAGCAAAACATTGCTCAAGCAGGAAGAGGACCTGGAACCATGACCATCCTCAGGGCTCAGATG GAAGTTTACTGTTCTGATTTTCATGCTGAAAGAGCAGGAGAGAGAAAATTCATGAGGAAAAG GAGCAACTGGCATTGCGAGCTGGCAGTTCTGCTGAAAGAGAATGATGCTTTTGAAGACGGAGGC AGGCAGTCCTTGATGGAGATGCAGAGTCGTATGGGGCGAGAACAGTGACTCTGACCAGCAG GCTTACCTTGTTCAAAGAGGAGCTGAGGACAGGACTGGCGGCAACAGCGGAATATTCCGATT CATTCTGCCCCAAGTGTGGAGAGGTTCTGCCTGACATAGACACGTTACAGATTACAGTGTATG GATTGCATCATTAAAGTGTTAATGTATCACCTCCCCAAAAGTGTGGTAAATGTGAGATTTTT TCCTCCAAGAGTTGTGCTTTTGTGTTATTTGTTTTCACTCAAATATTTTGCTCATTATTCTT GTTTTAAAGAAAGAAAACAGGCCGGGCACAGTGGCTCATGCCTGTAATCCAGCACTTTGGG AGATCCAGGTGGGAGGAT		
	ORF Start: ATG at 31		ORF Stop: TAA at 1714
	SEQ ID NO: 336	561 aa	MW at 64354.8kD
NOV91d, CG97090-02 Protein Sequence	MSHQPLRSSHLDPNLDFTPEELLQMKELLTENHQLKEAMKLNQAMKGRFEELSAWTEKQ KEERQFFEIQSKEAKERLMALSHENEKLEELGKLKGKSERSSDPTDSSLRAEAEQEKDQ LRTQVVRQLQAEKADLLGIVSELQKLKLNSSGSSSDSFVEIRMAEGEAEGSVKEIKHSPGPTRTV STGTALSKYRSRSADGAKNYFEHEELTVSQQLLCLREGNQKVERLEVALKEAKERVSDFEKKT SNRSEIETQTEGSTEKENDEEKGPEVTGSEVEALNLQVTSFLFKELQEAHTKLSEAEMLMKRLQ EKQALERKNSAIPSELNEKQELVYTNKKLELQVESMLSEIKMEQAKTEDEKSKLTVLQMTN KLLQEHNNALKTIEELTRKESEKVDRAVLKELSEKLELAEKALASKQLQMDMKQTIKQEEED LETMTILRAQMEVYCSDFHAERAAREKIHEEKEQLALQLAVLLKENDAFEDGGRQSLMEMQSR HGARTSDSDQAYLVQRGAEDRDWRQRNPIHSCPKCGEVLDPDITLQIHVMDCII		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 91B.

Table 91B. Comparison of NOV91a against NOV91b through NOV91d.		
Protein Sequence	NOV91a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV91b	34..561 1..534	494/534 (92%) 494/534 (92%)
NOV91c	34..561 1..528	494/528 (93%) 494/528 (93%)
NOV91d	1..561 1..561	520/567 (91%) 520/567 (91%)

5

Further analysis of the NOV91a protein yielded the following properties shown in Table 91C.

Table 91C. Protein Sequence Properties NOV91a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3000 probability located in microbody (peroxisome); 0.1000 probability located in mitochondrial matrix space; 0.1000 probability located in lysosome (lumen)

SignalP analysis:	No Known Signal Sequence Predicted
-------------------	------------------------------------

A search of the NOV91a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 91D.

5

Table 91D. Geneseq Results for NOV91a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV91a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AA Y27431	Murine RIP-associated protein (RAP-2) splice variant (NEMO full) - Mus sp, 412 aa. [WO9947672-A1, 23-SEP-1999]	232..560 90..412	100/341 (29%) 184/341 (53%)	3e-33
AA Y27430	Human RIP-associated protein (RAP-2) - Homo sapiens, 416 aa. [WO9947672-A1, 23-SEP-1999]	229..558 88..416	101/337 (29%) 184/337 (53%)	6e-32
AAU84350	Protein MYH11 differentially expressed in breast cancer tissue - Homo sapiens, 1857 aa. [WO200210436-A2, 07-FEB-2002]	29..502 1095..1596	117/520 (22%) 222/520 (42%)	1e-15
ABG06505	Novel human diagnostic protein #6496 - Homo sapiens, 2633 aa. [WO200175067-A2, 11-OCT-2001]	29..487 1054..1546	117/511 (22%) 230/511 (44%)	1e-15
AAM41000	Human polypeptide SEQ ID NO 5931 - Homo sapiens, 1988 aa. [WO200153312-A1, 26-JUL-2001]	22..502 983..1526	126/563 (22%) 236/563 (41%)	2e-15

In a BLAST search of public sequence databases, the NOV91a protein was found to have homology to the proteins shown in the BLASTP data in Table 91E.

Table 91E. Public BLASTP Results for NOV91a				
Protein Accession Number	Protein/Organism/Length	NOV91a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
AAH32762	Similar to optineurin - Homo sapiens (Human), 571 aa.	1..561 1..571	555/571 (97%) 556/571 (97%)	0.0

Q96CV9	Tumor necrosis factor alpha-inducible cellular protein containing leucine zipper domains, huntingtin interacting protein L, transcription factor IIIA-interacting protein (Optineurin isoform 1) (Optineurin isoform 2) (Optineurin isoform 3) - Homo sapiens (Human), 577 aa.	1..561 1..577	555/577 (96%) 556/577 (96%)	0.0
Q9Y218	FIP2 - Homo sapiens (Human), 577 aa.	1..561 1..577	552/577 (95%) 554/577 (95%)	0.0
Q9BGR3	Hypothetical 65.1 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 571 aa.	1..561 1..571	538/571 (94%) 547/571 (95%)	0.0
Q95KA2	Hypothetical 62.9 kDa protein - Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey), 550 aa.	16..561 5..550	526/546 (96%) 534/546 (97%)	0.0

PFam analysis predicts that the NOV91a protein contains the domains shown in the Table 91F.

Table 91F. Domain Analysis of NOV91a			
Pfam Domain	NOV91a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C2H2	537..559	6/24 (25%) 17/24 (71%)	0.51

5

Example 92.

The NOV92 clone was analyzed, and the nucleotide and encoded polypeptide sequences are shown in Table 92A.

Table 92A. NOV92 Sequence Analysis			
	SEQ ID NO: 337	750 bp	
NOV92a, CG97966-01 DNA Sequence	GTGGCTGCTCGGGACCAACCGACCCGCGGCCATGGCCCCGGCCCGCCAGCCCCCGGAGG TGATCCGCGCGGCGCAGAAAGGACGAGTACTACCGCGGTGGGCTGCGGAGCGCGGCGGGCGGCG CCCTGCACAGCCTGGCGGGTGCGGGGAAGTGGCTGGAGTGGAGGAAGGAGGTTGAGCTGCTCT CAGATGTGGCCTACTTTGGCCTCACCACACTTGCAGGCTACCAGACCCTGGGGGAGGAGTACG TCAGCATCATCCAGGTGGACCCATCGCGGATACATGTGCCCTCCTCGCTGCGCCGTGGCGTGC TGGTGACGCTGCATGCGTCTCCTGCCCTACCTGCTGGACAAGGCCCTGCTCCCCCTGGAGCAGG AGCTGCAGGCTGACCCCGACAGTGGGCGACCCCTTGCAGGGGAGCCTGGGGCCAGGTGGGCGTG GCTGCTCAGGGGCGCGGCGCTGGATGCGTCACCACACGGCCACCCTGACTGAGCAGCAGAGGA GGGCGCTGCTGCGGGCGGTCTTCGTCTCAGACAGGGCCTCGCCTGCCTCCAGCGGCTACATG TTGCCTGGTTTTACATCCACCTGTTCTGCTGGGAGTGCATCACCAGGTGGTGCAGCAGCAAGG CGGAGTGTCCCTCTGCGGGGAGAAGTTCCTCCCCAGAAGCTCATCTACCTTCGGCACTACC GCTGAGCCGCGCCCGGGTGGGCTGGACACAGATGACCTCTACGGGAGTCTGAACG		
	ORF Start: ATG at 33		ORF Stop: TGA at 696

	SEQ ID NO: 338	221 aa	MW at 24759.4kD
NOV92a, CG97966-01 Protein Sequence	MAPAAASPPEVIRAAQKDEYYRGGLRSAAGGALHSLAGAGKLEWRKEVELLSDVAYFGLTTL AGYQTLGEEYVSI IQVDPRIHVPSLRRLRVLT LHAVLPYLLDKALLPLEQELQADPDSGRP LQGS LGPGRGCSGARRWMRHHTATLTEQRRALLRAVFLRQGLACLQRLHVAFYIHLFCW ECITAWCSSKAECPLCREKFPQKLIYLRHYR		
	SEQ ID NO: 339	489 bp	
NOV92b, CG97966-03 DNA Sequence	GTGGCTGCTCGGGACCACCCGAACCCGCGGCCATGGCCCCGGCCGCCAGCCCCCGGAGG TGATCCGCGCGGCGCAGAAGGACGAGTACTACCGCGGTGGGCTGCGGAGCGCGGCGGGCGGCG CCCTGCACAGCCTGGCGGGTGCGAGGAAGTGGCTGGAGTGGAGGAAGGAGTTGAGCTGCTCT CAGATGTGGCCTACTTTGGCCTCACCACACTTGCAGGCTACCAGACCCTGGGGGAGGAGAGAG CCGTTTCCAGAAACCCCTGTGCACCCTGTGCTGGAGGAGCGCAGGCACCCAACAGCCACGC CCTGCGGCCACCTGTTCTGCTGGGAGTGCATCACCGCGTGGTGCAGCAGCAAGGCGGAGTGTG CCCTCTGCGGGAGAAGTTCCTCCCCAGAAGCTCATCTACCTTCGGCACTACCGCTGAGCCG GCGCCCGGGTGGCCTGGACACAGATGACCTCTACGGGAGTCTGAACG		
	ORF Start: ATG at 33		ORF Stop: TGA at 435
	SEQ ID NO: 340	134 aa	MW at 15069.1kD
NOV92b, CG97966-03 Protein Sequence	MAPAAASPPEVIRAAQKDEYYRGGLRSAAGGALHSLAGARKLEWRKEVELLSDVAYFGLTTL AGYQTLGEEERAVSRNPLCTLCLEERRHPTATPCGHLFCWECITAWCSSKAECPLCREKFPQK LIYLRHYR		
	SEQ ID NO: 341	1267 bp	
NOV92c, CG97966-02 DNA Sequence	GTGGCTGCTCGGGACCACCCGAACCCGCGGCCATGGCCCCGGCCGCCAGCCCCCGGAGG TGATCCGCGCGGCGCAGAAGGACGAGTACTACCGCGGTGGGCTGCGGAGCGCGGCGGGCGGCG CCCTGCACAGCCTGGCGGGTGCGAGGAAGTGGCTGGAGTGGAGGAAGGAGTTGAGCTGCTCT CAGATGTGGCCTACTTTGGCCTCACCACACTTGCAGGCTACCAGACCCTGGGGGAGGAGTACG TCAGCATCATCCAGGTGGACCCATCGCGGATACATGTGCCCTCCTCGCTGCGCCGTGGCGTGC TGGTGACGCTGCATGCCGTCTGCCCTACCTGCTGGACAAGGCCCTGCTCCCCCTGGAGCAGG AGCTGCAGGCTGACCCGACAGTGGGCGACCCTTGAGGGGAGCCTGGGGCCAGGTGGGCGTG GCTGCTCAGGGGCGCGCGCTGGATGCGTCACCACACGGCCACCCTGACTGAGCAGCAGAGGA GGGCGCTGCTGCGGGCGGTCTTCGTCTCAGACAGGGCCTCGCCTGCCTCCAGCGGTACATG TTGCCTGGTTTACATCCACGGTGTCTTCTACCACCTGGCCAAGAGGCTCAGGGGATCACGT ACCTCCGTGTCCGAGCCTGCCCGGAGAGGACCTGAGGGCCCGTGTAGCTACAGCTGCTGG GGTTCATCTCACTGCTGCACCTGGTGTGCTGTCATGGGGCTGCAGCTGTACGGTTTCAGGCAGC GGCAGCGAGCCAGGAAGGAGTGGAGGCTGCACCGCGCCTGTCTCACCAGGGCCCTCCTTGG AGGAGAGAGCCGTTTCCAGAAACCCCTGTGCACCCTGTGCTGGAGGAGCGCAGGCACCCAA CAGCCACGCCCTGCGGCCACCTGTTCTGCTGGGAGTGCATCACCGCGTGGTGCAGCAGCAAGG CGGAGTGTCCCTCTGCCGGGAGAAGTTCCTCCCCAGAAGCTCATCTACCTTCGGCACTACC GCTGAGCCCGGCGCCCGGTGGGCTGGACACAGATGACCTCTACGGGAGTCTAAACGCCAAGA TTTAGTCTCAGGATTAACTTGTCTGCACAGAAGTTAGAACACTCTCAGTTTTTTGTCATGTA AGATACTAACCTAGCCACCCTGGGAGAGAACAGAAAGCTGTCCCTGGCTGCGCTTCTCAGCC CTGGGAGGGGCGCCTGAACCCAGAACATTTCCCTAACCCCAACCTGGTAGGACTCAGCCACTT CTTCAGG		
	ORF Start: ATG at 33		ORF Stop: TGA at 1011
	SEQ ID NO: 342	326 aa	MW at 37068.6kD
NOV92c, CG97966-02 Protein Sequence	MAPAAASPPEVIRAAQKDEYYRGGLRSAAGGALHSLAGARKLEWRKEVELLSDVAYFGLTTL AGYQTLGEEYVSI IQVDPRIHVPSLRRLRVLT LHAVLPYLLDKALLPLEQELQADPDSGRP LQGS LGPGRGCSGARRWMRHHTATLTEQRRALLRAVFLRQGLACLQRLHVAFYIHGVFY HLAKRLTGITYLRVRS LPPGDLRARVSYRL LGVISLLHLVLSMGLQLYGFRQRQRARKEWRLH RGLSHRRASLEERAVSRNPLCTLCLEERRHPTATPCGHLFCWECITAWCSSKAECPLCREKFP PQKLIYLRHYR		
	SEQ ID NO: 343	2059 bp	
NOV92d, CG97966-04 DNA Sequence	GGCAGGAGGTGCTCGGGACCACCCGAACCCGCGGCCATGGCCCCGGCCGCCAGCCCCCGGAGG GGAGGTGATCCGCGCGGCGCAGAAGGACGAGTACTACCGCGGTGGGCTGCGGAGCGCGGCGGG CGGCGCCCTGCACAGCCTGGCGGGTGCGAGGAAGTGGCTGGAGTGGAGGAAGGAGTTGAGCT GCTCTCAGATGTGGCCTACTTTGGCCTCACCACACTTGCAGGCTACCAGACCCTGGGGGAGGA GTACGTGAGCATCATCCAGGTGGACCCATCGCGGATACATGTGCCCTCCTCGCTGCGCCGTGG CGTGCTGGTGACGCTGCATGCCGTCTGCCCTACCTGCTGGACAAGGCCCTGCTCCCCCTGGA GCAGGAGCTGCAGGCTGACCCGACAGTGGGCGACCCTTGAGGGGAGCCTGGGGCCAGGTGG GCGTGGCTGCTCAGGGGCGCGCGCTGGATGCGTCACCACACGGCCACCCTGACTGAGCAGCA GAGGAGGGCGCTGCTGCGGGCGGTCTTCGTCTCAGACAGGGCCTCGCCTGCCTCCAGCGGCT ACATGTTGCTGGTTTACATCCACGGTGTCTTCTACCACCTGGCCAAGAGGCTCAGCGGGAT		

	CACGTACGCGCTGAGGCCAGATCCCTCAGGGTCCTGATGAGTGTGGCGCCATCTGCCTTACA GCTCCGTGTCCGAGCCTGCCCGAGAGGACCTGAGGGCCCGTGTAGTACAGGCTGTGGG GGTCATCTCACTGCTGCACCTGGTGTGTCCATGGGGCTGCAGCTGTACGGTTTCAGGCAGCG GCAGCGAGCCAGGAAGGAGTGGAGGCTGCACCGCGGCTGTCTACCCGAGGGCCTCCTTGA GGAGAGAGCCGTTTCCAGAAACCCCTGTGCACCCTGTGCCTGGAGGAGCGCAGGCACCCAAC AGCCACGCCCTGCGGCCACCTGTTCTGCTGGGAGTGCATCACCGCGTGGTGCAGCAGCAAGGC GGAGTGTCCCTCTGCCGGGAGAAGTTCCTCCCCAGAAGCTCATCTACCTTCGGCACTACCG CTGAGCCGGCGCCCGGGTGGGCCTGGACACAGATGACCTCTACGGGAGTCTGAACGCCAAGAT TTAGTCTCAGGATTAACCTTGCTTGACAGAAAGTTAGAACACTCTCAGTTTTTGTCTATGTA GATACTAACCTAGCCACCCTGGGAGAGAACAGAAAGCTGTCCCTGGCTGCACTTCTCAGCCC TGGGAGGGGCGCCTGAACCCAGAACATTTCCCTAACCCCAACCTGGTAGGACTCAGCCACTTC TTCAGGAATTTCACTTATTTGGACGGGATTTAGGTTTCCCTCCCTTCCCCAACCATACAGT TGAGAAGTAATTCAGAAGTAGGCCAGAAGACACTTTATTTCGTTTATATTGTGAGAAAACAGCC CCATCAGGCTTGTGTTAAGGCAATGGACTGAATGAGTGCCTGCTGGGTGGGGTGGGGCAGGA GGCTGGCGGGTGTCTCAGCCAGTGCAGTGAGAACAGCAGCCCCACGGCCCCATGGGAGGCGG CGCTGCTCTCCCCGAGGGCGGCTGGGCAGAGCACATCCCCAGGACTTGATGACCACACGGGG CAGAGAGAAACCAACCAAGGCCAGCACCTCCGTCGGAAGCATTGGGCACACACCTTCAATA CACGTCAAGGTCGCTTCCAGTTTTAGAAAACAGAAATCTGCATCTCAGCCTGAGACGCACAGA GAGGTCTCTTCTGACCCAGACGCACTCACGAGCCAGGTCTGGGGGTATGGGGGCTGCCAGG GGCGCCGAGCCCTCTCTGGGGGGCTGTGGGCAGGCGACCTGCTGACCCACGGTCACTGC TGTGTTACGCCCCCTCAGCTCGGCCCCAGCCTATTTCCCGCTCCATTGATGTTTCCAGGTTT TCAAACTGCATTAACTGCGCCAGAGAGTTCACCGTAGGCATCTTAAATAAACTAACTCCA GCAAAATGTGGGTACGTTACTAAAAA		
	ORF Start: ATG at 38		ORF Stop: TGA at 1073
	SEQ ID NO: 344	345 aa	MW at 39085.0kD
NOV92d, CG97966-04 Protein Sequence	MAPAAASPPEVIRAAQKDEYYRGLRSAAGGALHSLAGARKWLEWRKEVELLSDVAYFGLTTL AGYQTLGEEYVSI IQVDPRIHVPSSLRRGVLVTLHAVLPYLLDKALLPLEQELQADPDSDGRP LQGS LGPGRGCSGARRWMRHHTATLTEQRRALLRAVFVLRQGLACLQRLHVAWFIYHGVFY HLAKRLTGITYALRPDPLRVLMVSAPSAQLRVRLPGEDLRARVSYRLGLVLSLLHLVLSMG LQLYGFRRQRARKEWRLHRGLSHRRASLEERAVSRNPLCTLCLEERRHPTATPCGHLFCWEC ITAWCSSKAECPLCREKFPQKLIYLRHYR		

Sequence comparison of the above protein sequences yields the following sequence relationships shown in Table 92B.

Table 92B. Comparison of NOV92a against NOV92b through NOV92d.		
Protein Sequence	NOV92a Residues/ Match Residues	Identities/ Similarities for the Matched Region
NOV92b	1..72 1..72	53/72 (73%) 53/72 (73%)
NOV92c	1..185 1..185	125/185 (67%) 125/185 (67%)
NOV92d	1..185 1..185	125/185 (67%) 125/185 (67%)

5

Further analysis of the NOV92a protein yielded the following properties shown in Table 92C.

Table 92C. Protein Sequence Properties NOV92a	
PSort analysis:	0.4500 probability located in cytoplasm; 0.3774 probability located in microbody (peroxisome); 0.2542 probability located in lysosome (lumen); 0.1000 probability located in mitochondrial matrix space
SignalP analysis:	No Known Signal Sequence Predicted

A search of the NOV92a protein against the Geneseq database, a proprietary database that contains sequences published in patents and patent publication, yielded several homologous proteins shown in Table 92D.

5

Table 92D. Geneseq Results for NOV92a				
Geneseq Identifier	Protein/Organism/Length [Patent #, Date]	NOV92a Residues/ Match Residues	Identities/ Similarities for the Matched Region	Expect Value
AAB51471	Human secreted protein BLAST search protein SEQ ID NO: 148 - Homo sapiens, 55 aa. [WO200058495-A1, 05-OCT-2000]	185..221 19..55	37/37 (100%) 37/37 (100%)	3e-18
AAB51470	Human secreted protein BLAST search protein SEQ ID NO: 147 - Homo sapiens, 55 aa. [WO200058495-A1, 05-OCT-2000]	185..221 19..55	37/37 (100%) 37/37 (100%)	3e-18
AAB51469	Human secreted protein BLAST search protein SEQ ID NO: 146 - Homo sapiens, 55 aa. [WO200058495-A1, 05-OCT-2000]	185..221 19..55	37/37 (100%) 37/37 (100%)	3e-18
AAB51468	Human secreted protein BLAST search protein SEQ ID NO: 145 - Homo sapiens, 55 aa. [WO200058495-A1, 05-OCT-2000]	185..221 19..55	37/37 (100%) 37/37 (100%)	3e-18
AAU93078	Arabidopsis transcription factor #116 - Arabidopsis thaliana, 381 aa. [WO200215675-A1, 28-FEB-2002]	6..108 29..131	39/103 (37%) 59/103 (56%)	6e-13

In a BLAST search of public sequence databases, the NOV92a protein was found to have homology to the proteins shown in the BLASTP data in Table 92E.

Table 92E. Public BLASTP Results for NOV92a				
Protein Accession Number	Protein/Organism/Length	NOV92a Residues/ Match Residues	Identities/ Similarities for the Matched Portion	Expect Value
O60683	Peroxisome assembly protein 10 (Peroxin-10) - Homo sapiens (Human), 326 aa.	1..185 1..185	184/185 (99%) 184/185 (99%)	e-103
AAM64667	Putative peroxisome assembly protein PER8 - Arabidopsis thaliana (Mouse-ear cress), 381 aa.	6..108 29..131	39/103 (37%) 59/103 (56%)	2e-12
Q9SYU4	Zinc-binding peroxisomal integral membrane protein (Putative peroxisome assembly protein PER8) - Arabidopsis thaliana (Mouse-ear cress), 381 aa.	6..108 29..131	39/103 (37%) 59/103 (56%)	2e-12
Q9M400	Pex10p - Arabidopsis thaliana (Mouse-ear cress), 381 aa.	6..108 29..131	39/103 (37%) 59/103 (56%)	2e-12
Q94LL6	Putative zinc-binding peroxisomal integral membrane protein - Oryza sativa (Rice), 382 aa.	6..108 31..133	38/103 (36%) 57/103 (54%)	4e-11

PFam analysis predicts that the NOV92a protein contains the domains shown in the Table 92F.

Table 92F. Domain Analysis of NOV92a			
Pfam Domain	NOV92a Match Region	Identities/ Similarities for the Matched Region	Expect Value
zf-C3HC4	185..205	9/29 (31%) 16/29 (55%)	0.011

5

Example B: Sequencing Methodology and Identification of NOVX Clones

1. **GeneCalling™ Technology:** This is a proprietary method of performing differential gene expression profiling between two or more samples developed at CuraGen and described by Shimkets, et al., "Gene expression analysis by transcript profiling coupled to a gene database query" Nature Biotechnology 17:198-803 (1999). cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained

as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then digested with up to as many as 120 pairs of restriction enzymes and pairs of linker-adaptors specific for each pair of restriction enzymes were ligated to the appropriate end. The restriction digestion generates a mixture of unique cDNA gene fragments. Limited PCR amplification is performed with primers homologous to the linker adapter sequence where one primer is biotinylated and the other is fluorescently labeled. The doubly labeled material is isolated and the fluorescently labeled single strand is resolved by capillary gel electrophoresis. A computer algorithm compares the electropherograms from an experimental and control group for each of the restriction digestions. This and additional sequence-derived information is used to predict the identity of each differentially expressed gene fragment using a variety of genetic databases. The identity of the gene fragment is confirmed by additional, gene-specific competitive PCR or by isolation and sequencing of the gene fragment.

2. **SeqCalling™ Technology:** cDNA was derived from various human samples representing multiple tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then sequenced using CuraGen's proprietary SeqCalling technology. Sequence traces were evaluated manually and edited for corrections if appropriate. cDNA sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

3. **PathCalling™ Technology:** The NOVX nucleic acid sequences are derived by laboratory screening of cDNA library by the two-hybrid approach. cDNA fragments covering either the full length of the DNA sequence, or part of the sequence, or both, are sequenced. In silico prediction was based on sequences available in CuraGen Corporation's

proprietary sequence databases or in the public human sequence databases, and provided either the full length DNA sequence, or some portion thereof.

The laboratory screening was performed using the methods summarized below:

cDNA libraries were derived from various human samples representing multiple
5 tissue types, normal and diseased states, physiological states, and developmental states from different donors. Samples were obtained as whole tissue, primary cells or tissue cultured primary cells or cell lines. Cells and cell lines may have been treated with biological or chemical agents that regulate gene expression, for example, growth factors, chemokines or steroids. The cDNA thus derived was then directionally cloned into the appropriate
10 two-hybrid vector (Gal4-activation domain (Gal4-AD) fusion). Such cDNA libraries as well as commercially available cDNA libraries from Clontech (Palo Alto, CA) were then transferred from E.coli into a CuraGen Corporation proprietary yeast strain (disclosed in U. S. Patents 6,057,101 and 6,083,693, incorporated herein by reference in their entireties).

Gal4-binding domain (Gal4-BD) fusions of a CuraGen Corporation proprietary library
15 of human sequences was used to screen multiple Gal4-AD fusion cDNA libraries resulting in the selection of yeast hybrid diploids in each of which the Gal4-AD fusion contains an individual cDNA. Each sample was amplified using the polymerase chain reaction (PCR) using non-specific primers at the cDNA insert boundaries. Such PCR product was sequenced; sequence traces were evaluated manually and edited for corrections if appropriate. cDNA
20 sequences from all samples were assembled together, sometimes including public human sequences, using bioinformatic programs to produce a consensus sequence for each assembly. Each assembly is included in CuraGen Corporation's database. Sequences were included as components for assembly when the extent of identity with another component was at least 95% over 50 bp. Each assembly represents a gene or portion thereof and includes information
25 on variants, such as splice forms single nucleotide polymorphisms (SNPs), insertions, deletions and other sequence variations.

Physical clone: the cDNA fragment derived by the screening procedure, covering the entire open reading frame is, as a recombinant DNA, cloned into pACT2 plasmid (Clontech) used to make the cDNA library. The recombinant plasmid is inserted into the host and
30 selected by the yeast hybrid diploid generated during the screening procedure by the mating of both CuraGen Corporation proprietary yeast strains N106' and YULH (U. S. Patents 6,057,101 and 6,083,693).

4. **RACE:** Techniques based on the polymerase chain reaction such as rapid amplification of cDNA ends (RACE), were used to isolate or complete the predicted sequence of the cDNA of the invention. Usually multiple clones were sequenced from one or more human samples to derive the sequences for fragments. Various human tissue samples from different donors were used for the RACE reaction. The sequences derived from these procedures were included in the SeqCalling Assembly process described in preceding paragraphs.

5. **Exon Linking:** The NOVX target sequences identified in the present invention were subjected to the exon linking process to confirm the sequence. PCR primers were designed by starting at the most upstream sequence available, for the forward primer, and at the most downstream sequence available for the reverse primer. In each case, the sequence was examined, walking inward from the respective termini toward the coding sequence, until a suitable sequence that is either unique or highly selective was encountered, or, in the case of the reverse primer, until the stop codon was reached. Such primers were designed based on in silico predictions for the full length cDNA, part (one or more exons) of the DNA or protein sequence of the target sequence, or by translated homology of the predicted exons to closely related human sequences from other species. These primers were then employed in PCR amplification based on the following pool of human cDNAs: adrenal gland, bone marrow, brain - amygdala, brain - cerebellum, brain - hippocampus, brain - substantia nigra, brain - thalamus, brain -whole, fetal brain, fetal kidney, fetal liver, fetal lung, heart, kidney, lymphoma - Raji, mammary gland, pancreas, pituitary gland, placenta, prostate, salivary gland, skeletal muscle, small intestine, spinal cord, spleen, stomach, testis, thyroid, trachea, uterus. Usually the resulting amplicons were gel purified, cloned and sequenced to high redundancy. The PCR product derived from exon linking was cloned into the pCR2.1 vector from Invitrogen. The resulting bacterial clone has an insert covering the entire open reading frame cloned into the pCR2.1 vector. The resulting sequences from all clones were assembled with themselves, with other fragments in CuraGen Corporation's database and with public ESTs. Fragments and ESTs were included as components for an assembly when the extent of their identity with another component of the assembly was at least 95% over 50 bp. In addition, sequence traces were evaluated manually and edited for corrections if appropriate. These procedures provide the sequence reported herein.

6. **Physical Clone:** Exons were predicted by homology and the intron/exon boundaries were determined using standard genetic rules. Exons were further selected and

refined by means of similarity determination using multiple BLAST (for example, tBlastN, BlastX, and BlastN) searches, and, in some instances, GeneScan and Grail. Expressed sequences from both public and proprietary databases were also added when available to further define and complete the gene sequence. The DNA sequence was then manually
5 corrected for apparent inconsistencies thereby obtaining the sequences encoding the full-length protein.

The PCR product derived by exon linking, covering the entire open reading frame, was cloned into the pCR2.1 vector from Invitrogen to provide clones used for expression and screening purposes.

10 **Example C: Quantitative expression analysis of clones in various cells and tissues**

The quantitative expression of various clones was assessed using microtiter plates containing RNA samples from a variety of normal and pathology-derived cells, cell lines and tissues using real time quantitative PCR (RTQ PCR). RTQ PCR was performed on an Applied Biosystems ABI PRISM® 7700 or an ABI PRISM® 7900 HT Sequence Detection
15 System. Various collections of samples are assembled on the plates, and referred to as Panel 1 (containing normal tissues and cancer cell lines), Panel 2 (containing samples derived from tissues from normal and cancer sources), Panel 3 (containing cancer cell lines), Panel 4 (containing cells and cell lines from normal tissues and cells related to inflammatory conditions), Panel 5D/5I (containing human tissues and cell lines with an emphasis on
20 metabolic diseases), AI_comprehensive_panel (containing normal tissue and samples from autoimmune/autoinflammatory diseases), Panel CNSD.01 (containing samples from normal and diseased brains) and CNS_neurodegeneration_panel (containing samples from normal and Alzheimer's diseased brains).

RNA integrity from all samples is controlled for quality by visual assessment of
25 agarose gel electropherograms using 28S and 18S ribosomal RNA staining intensity ratio as a guide (2:1 to 2.5:1 28s:18s) and the absence of low molecular weight RNAs that would be indicative of degradation products. Samples are controlled against genomic DNA contamination by RTQ PCR reactions run in the absence of reverse transcriptase using probe and primer sets designed to amplify across the span of a single exon.

30 First, the RNA samples were normalized to reference nucleic acids such as constitutively expressed genes (for example, β -actin and GAPDH). Normalized RNA (5 μ l) was converted to cDNA and analyzed by RTQ-PCR using One Step RT-PCR Master Mix

Reagents (Applied Biosystems; Catalog No. 4309169) and gene-specific primers according to the manufacturer's instructions.

In other cases, non-normalized RNA samples were converted to single strand cDNA (sscDNA) using Superscript II (Invitrogen Corporation; Catalog No. 18064-147) and random
5 hexamers according to the manufacturer's instructions. Reactions containing up to 10 µg of total RNA were performed in a volume of 20 µl and incubated for 60 minutes at 42 °C. This reaction can be scaled up to 50 µg of total RNA in a final volume of 100 µl. sscDNA samples are then normalized to reference nucleic acids as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the
10 manufacturer's instructions.

Probes and primers were designed for each assay according to Applied Biosystems Primer Express Software package (version I for Apple Computer's Macintosh Power PC) or a similar algorithm using the target sequence as input. Default settings were used for reaction conditions and the following parameters were set before selecting primers: primer
15 concentration = 250 nM, primer melting temperature (T_m) range = 58 °-60 °C, primer optimal T_m = 59 °C, maximum primer difference = 2 °C, probe does not have 5'G, probe T_m must be 10 °C greater than primer T_m , amplicon size 75bp to 100bp. The probes and primers selected (see below) were synthesized by Synthegen (Houston, TX, USA). Probes were double
20 purified by HPLC to remove uncoupled dye and evaluated by mass spectroscopy to verify coupling of reporter and quencher dyes to the 5' and 3' ends of the probe, respectively. Their final concentrations were: forward and reverse primers, 900nM each, and probe, 200nM.

PCR conditions: When working with RNA samples, normalized RNA from each tissue and each cell line was spotted in each well of either a 96 well or a 384-well PCR plate (Applied Biosystems). PCR cocktails included either a single gene specific probe and primers
25 set, or two multiplexed probe and primers sets (a set specific for the target clone and another gene-specific set multiplexed with the target probe). PCR reactions were set up using TaqMan® One-Step RT-PCR Master Mix (Applied Biosystems, Catalog No. 4313803) following manufacturer's instructions. Reverse transcription was performed at 48°C for 30 minutes followed by amplification/PCR cycles as follows: 95°C 10 min, then 40 cycles of 95
30 °C for 15 seconds, 60 °C for 1 minute. Results were recorded as CT values (cycle at which a given sample crosses a threshold level of fluorescence) using a log scale, with the difference in RNA concentration between a given sample and the sample with the lowest CT value

being represented as 2 to the power of delta CT. The percent relative expression is then obtained by taking the reciprocal of this RNA difference and multiplying by 100.

When working with sscDNA samples, normalized sscDNA was used as described previously for RNA samples. PCR reactions containing one or two sets of probe and primers were set up as described previously, using 1X TaqMan® Universal Master mix (Applied Biosystems; catalog No. 4324020), following the manufacturer's instructions. PCR amplification was performed as follows: 95 °C 10 min, then 40 cycles of 95 °C for 15 seconds, 60 °C for 1 minute. Results were analyzed and processed as described previously.

Panels 1, 1.1, 1.2, and 1.3D

The plates for Panels 1, 1.1, 1.2 and 1.3D include 2 control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in these panels are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in these panels are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal tissues found on these panels are comprised of samples derived from all major organ systems from single adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea, breast, ovary, uterus, placenta, prostate, testis and adipose.

In the results for Panels 1, 1.1, 1.2 and 1.3D, the following abbreviations are used:

ca. = carcinoma,

* = established from metastasis,

met = metastasis,

s cell var = small cell variant,

non-s = non-sm = non-small,

squam = squamous,

pl. eff = pl effusion = pleural effusion,
glio = glioma,
astro = astrocytoma, and
neuro = neuroblastoma.

5 **General_screening_panel_v1.4, v1.5 and v1.6**

The plates for Panels 1.4, v1.5 and v1.6 include two control wells (genomic DNA control and chemistry control) and 94 wells containing cDNA from various samples. The samples in Panels 1.4, v1.5 and v1.6 are broken into 2 classes: samples derived from cultured cell lines and samples derived from primary normal tissues. The cell lines are derived from
10 cancers of the following types: lung cancer, breast cancer, melanoma, colon cancer, prostate cancer, CNS cancer, squamous cell carcinoma, ovarian cancer, liver cancer, renal cancer, gastric cancer and pancreatic cancer. Cell lines used in Panels 1.4, v1.5 and v1.6 are widely available through the American Type Culture Collection (ATCC), a repository for cultured cell lines, and were cultured using the conditions recommended by the ATCC. The normal
15 tissues found on Panels 1.4, v1.5 and v1.6 are comprised of pools of samples derived from all major organ systems from 2 to 5 different adult individuals or fetuses. These samples are derived from the following organs: adult skeletal muscle, fetal skeletal muscle, adult heart, fetal heart, adult kidney, fetal kidney, adult liver, fetal liver, adult lung, fetal lung, various regions of the brain, the spleen, bone marrow, lymph node, pancreas, salivary gland, pituitary
20 gland, adrenal gland, spinal cord, thymus, stomach, small intestine, colon, bladder, trachea. breast, ovary, uterus, placenta, prostate, testis and adipose. Abbreviations are as described for Panels 1, 1.1, 1.2, and 1.3D.

Panels 2D, 2.2, 2.3 and 2.4

The plates for Panels 2D, 2.2, 2.3 and 2.4 generally include two control wells and 94
25 test samples composed of RNA or cDNA isolated from human tissue procured by surgeons working in close cooperation with the National Cancer Institute's Cooperative Human Tissue Network (CHTN) or the National Disease Research Initiative (NDRI) or from Ardaïs or Clinomics. The tissues are derived from human malignancies and in cases where indicated many malignant tissues have "matched margins" obtained from noncancerous tissue just
30 adjacent to the tumor. These are termed normal adjacent tissues and are denoted "NAT" in the results below. The tumor tissue and the "matched margins" are evaluated by two independent pathologists (the surgical pathologists and again by a pathologist at NDRI/

CHTN/Ardais/Clinomics). Unmatched RNA samples from tissues without malignancy (normal tissues) were also obtained from Ardais or Clinomics. This analysis provides a gross histopathological assessment of tumor differentiation grade. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical stage of the patient. These matched margins are taken from the tissue surrounding (*i.e.* immediately proximal) to the zone of surgery (designated "NAT", for normal adjacent tissue, in Table RR). In addition, RNA and cDNA samples were obtained from various human tissues derived from autopsies performed on elderly people or sudden death victims (accidents, *etc.*). These tissues were ascertained to be free of disease and were purchased from various commercial sources such as Clontech (Palo Alto, CA), Research Genetics, and Invitrogen. General oncology screening panel_v_2.4 is an updated version of Panel 2D.

HASS Panel v 1.0

The HASS panel v 1.0 plates are comprised of 93 cDNA samples and two controls. Specifically, 81 of these samples are derived from cultured human cancer cell lines that had been subjected to serum starvation, acidosis and anoxia for different time periods as well as controls for these treatments, 3 samples of human primary cells, 9 samples of malignant brain cancer (4 medulloblastomas and 5 glioblastomas) and 2 controls. The human cancer cell lines are obtained from ATCC (American Type Culture Collection) and fall into the following tissue groups: breast cancer, prostate cancer, bladder carcinomas, pancreatic cancers and CNS cancer cell lines. These cancer cells are all cultured under standard recommended conditions. The treatments used (serum starvation, acidosis and anoxia) have been previously published in the scientific literature. The primary human cells were obtained from Clonetics (Walkersville, MD) and were grown in the media and conditions recommended by Clonetics. The malignant brain cancer samples are obtained as part of a collaboration (Henry Ford Cancer Center) and are evaluated by a pathologist prior to CuraGen receiving the samples. RNA was prepared from these samples using the standard procedures. The genomic and chemistry control wells have been described previously.

ARDAIS Panel v 1.0

The plates for ARDAIS panel v 1.0 generally include 2 control wells and 22 test samples composed of RNA isolated from human tissue procured by surgeons working in close cooperation with Ardais Corporation. The tissues are derived from human lung malignancies (lung adenocarcinoma or lung squamous cell carcinoma) and in cases where

indicated many malignant samples have “matched margins” obtained from noncancerous lung tissue just adjacent to the tumor. These matched margins are taken from the tissue surrounding (i.e. immediately proximal) to the zone of surgery (designated “NAT”, for normal adjacent tissue) in the results below. The tumor tissue and the “matched margins” are evaluated by independent pathologists (the surgical pathologists and again by a pathologist at Ardaïs). Unmatched malignant and non-malignant RNA samples from lungs were also obtained from Ardaïs. Additional information from Ardaïs provides a gross histopathological assessment of tumor differentiation grade and stage. Moreover, most samples include the original surgical pathology report that provides information regarding the clinical state of the patient.

Panels 3D and 3.1

The plates of Panels 3D and 3.1 are comprised of 94 cDNA samples and two control samples. Specifically, 92 of these samples are derived from cultured human cancer cell lines, 2 samples of human primary cerebellar tissue and 2 controls. The human cell lines are generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups: Squamous cell carcinoma of the tongue, breast cancer, prostate cancer, melanoma, epidermoid carcinoma, sarcomas, bladder carcinomas, pancreatic cancers, kidney cancers, leukemias/lymphomas, ovarian/uterine/cervical, gastric, colon, lung and CNS cancer cell lines. In addition, there are two independent samples of cerebellum. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. The cell lines in panel 3D and 1.3D are of the most common cell lines used in the scientific literature. Oncology_cell_line_screening_panel_v3.2 is an updated version of Panel 3. The cell lines in panel 3D, 3.1, 1.3D and oncology_cell_line_screening_panel_v3.2 are of the most common cell lines used in the scientific literature.

Panels 4D, 4R, and 4.1D

Panel 4 includes samples on a 96 well plate (2 control wells, 94 test samples) composed of RNA (Panel 4R) or cDNA (Panels 4D/4.1D) isolated from various human cell lines or tissues related to inflammatory conditions. Total RNA from control normal tissues such as colon and lung (Stratagene, La Jolla, CA) and thymus and kidney (Clontech) was employed. Total RNA from liver tissue from cirrhosis patients and kidney from lupus patients was obtained from BioChain (Biochain Institute, Inc., Hayward, CA). Intestinal tissue for

RNA preparation from patients diagnosed as having Crohn's disease and ulcerative colitis was obtained from the National Disease Research Interchange (NDRI) (Philadelphia, PA).

Astrocytes, lung fibroblasts, dermal fibroblasts, coronary artery smooth muscle cells, small airway epithelium, bronchial epithelium, microvascular dermal endothelial cells, microvascular lung endothelial cells, human pulmonary aortic endothelial cells, human umbilical vein endothelial cells were all purchased from Clonetics (Walkersville, MD) and grown in the media supplied for these cell types by Clonetics. These primary cell types were activated with various cytokines or combinations of cytokines for 6 and/or 12-14 hours, as indicated. The following cytokines were used; IL-1 beta at approximately 1-5ng/ml, TNF alpha at approximately 5-10ng/ml, IFN gamma at approximately 20-50ng/ml, IL-4 at approximately 5-10ng/ml, IL-9 at approximately 5-10ng/ml, IL-13 at approximately 5-10ng/ml. Endothelial cells were sometimes starved for various times by culture in the basal media from Clonetics with 0.1% serum.

Mononuclear cells were prepared from blood of employees at CuraGen Corporation, using Ficoll. LAK cells were prepared from these cells by culture in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco/Life Technologies, Rockville, MD), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) and Interleukin 2 for 4-6 days. Cells were then either activated with 10-20ng/ml PMA and 1-2µg/ml ionomycin, IL-12 at 5-10ng/ml, IFN gamma at 20-50ng/ml and IL-18 at 5-10ng/ml for 6 hours. In some cases, mononuclear cells were cultured for 4-5 days in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco) with PHA (phytohemagglutinin) or PWM (pokeweed mitogen) at approximately 5µg/ml. Samples were taken at 24, 48 and 72 hours for RNA preparation. MLR (mixed lymphocyte reaction) samples were obtained by taking blood from two donors, isolating the mononuclear cells using Ficoll and mixing the isolated mononuclear cells 1:1 at a final concentration of approximately 2×10^6 cells/ml in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol (5.5×10^{-5} M) (Gibco), and 10mM Hepes (Gibco). The MLR was cultured and samples taken at various time points ranging from 1- 7 days for RNA preparation.

Monocytes were isolated from mononuclear cells using CD14 Miltenyi Beads, +ve VS selection columns and a Vario Magnet according to the manufacturer's instructions. Monocytes were differentiated into dendritic cells by culture in DMEM 5% fetal calf serum

(FCS) (Hyclone, Logan, UT), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco), 50ng/ml GMCSF and 5ng/ml IL-4 for 5-7 days. Macrophages were prepared by culture of monocytes for 5-7 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), 10mM Hepes (Gibco) and 10% AB Human Serum or MCSF at approximately 50ng/ml. Monocytes, macrophages and dendritic cells were stimulated for 6 and 12-14 hours with lipopolysaccharide (LPS) at 100ng/ml. Dendritic cells were also stimulated with anti-CD40 monoclonal antibody (Pharmingen) at 10 μ g/ml for 6 and 12-14 hours.

- 10 CD4 lymphocytes, CD8 lymphocytes and NK cells were also isolated from mononuclear cells using CD4, CD8 and CD56 Miltenyi beads, positive VS selection columns and a Vario Magnet according to the manufacturer's instructions. CD45RA and CD45RO CD4 lymphocytes were isolated by depleting mononuclear cells of CD8, CD56, CD14 and CD19 cells using CD8, CD56, CD14 and CD19 Miltenyi beads and positive selection.
- 15 CD45RO beads were then used to isolate the CD45RO CD4 lymphocytes with the remaining cells being CD45RA CD4 lymphocytes. CD45RA CD4, CD45RO CD4 and CD8 lymphocytes were placed in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and plated at 10⁶ cells/ml onto Falcon 6 well tissue culture plates that had been
- 20 coated overnight with 0.5 μ g/ml anti-CD28 (Pharmingen) and 3 μ g/ml anti-CD3 (OKT3, ATCC) in PBS. After 6 and 24 hours, the cells were harvested for RNA preparation. To prepare chronically activated CD8 lymphocytes, we activated the isolated CD8 lymphocytes for 4 days on anti-CD28 and anti-CD3 coated plates and then harvested the cells and expanded them in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco),
- 25 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and IL-2. The expanded CD8 cells were then activated again with plate bound anti-CD3 and anti-CD28 for 4 days and expanded as before. RNA was isolated 6 and 24 hours after the second activation and after 4 days of the second expansion culture. The isolated NK cells were cultured in DMEM 5% FCS (Hyclone), 100 μ M non essential amino
- 30 acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5x10⁻⁵M (Gibco), and 10mM Hepes (Gibco) and IL-2 for 4-6 days before RNA was prepared.

To obtain B cells, tonsils were procured from NDRI. The tonsil was cut up with sterile dissecting scissors and then passed through a sieve. Tonsil cells were then spun down

and resuspended at 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco). To activate the cells, we used PWM at 5 μ g/ml or anti-CD40 (Pharmingen) at approximately 10 μ g/ml and IL-4 at 5-10ng/ml. Cells were harvested for
5 RNA preparation at 24, 48 and 72 hours.

To prepare the primary and secondary Th1/Th2 and Tr1 cells, six-well Falcon plates were coated overnight with 10 μ g/ml anti-CD28 (Pharmingen) and 2 μ g/ml OKT3 (ATCC), and then washed twice with PBS. Umbilical cord blood CD4 lymphocytes (Poietic Systems, German Town, MD) were cultured at 10^5 - 10^6 cells/ml in DMEM 5% FCS (Hyclone), 100 μ M
10 non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco) and IL-2 (4ng/ml). IL-12 (5ng/ml) and anti-IL4 (1 μ g/ml) were used to direct to Th1, while IL-4 (5ng/ml) and anti-IFN gamma (1 μ g/ml) were used to direct to Th2 and IL-10 at 5ng/ml was used to direct to Tr1. After 4-5 days, the activated Th1, Th2 and Tr1 lymphocytes were washed once in DMEM and expanded for 4-7
15 days in DMEM 5% FCS (Hyclone), 100 μ M non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco) and IL-2 (1ng/ml). Following this, the activated Th1, Th2 and Tr1 lymphocytes were re-stimulated for 5 days with anti-CD28/OKT3 and cytokines as described above, but with the addition of anti-CD95L (1 μ g/ml) to prevent apoptosis. After 4-5 days, the Th1, Th2 and Tr1
20 lymphocytes were washed and then expanded again with IL-2 for 4-7 days. Activated Th1 and Th2 lymphocytes were maintained in this way for a maximum of three cycles. RNA was prepared from primary and secondary Th1, Th2 and Tr1 after 6 and 24 hours following the second and third activations with plate bound anti-CD3 and anti-CD28 mAbs and 4 days into the second and third expansion cultures in Interleukin 2.

25 The following leukocyte cells lines were obtained from the ATCC: Ramos, EOL-1, KU-812. EOL cells were further differentiated by culture in 0.1mM dbcAMP at 5×10^5 cells/ml for 8 days, changing the media every 3 days and adjusting the cell concentration to 5×10^5 cells/ml. For the culture of these cells, we used DMEM or RPMI (as recommended by the ATCC), with the addition of 5% FCS (Hyclone), 100 μ M non essential
30 amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), 10mM Hepes (Gibco). RNA was either prepared from resting cells or cells activated with PMA at 10ng/ml and ionomycin at 1 μ g/ml for 6 and 14 hours. Keratinocyte line CCD106 and an airway epithelial tumor line NCI-H292 were also obtained from the ATCC. Both were

cultured in DMEM 5% FCS (Hyclone), 100µM non essential amino acids (Gibco), 1mM sodium pyruvate (Gibco), mercaptoethanol 5.5×10^{-5} M (Gibco), and 10mM Hepes (Gibco). CCD1106 cells were activated for 6 and 14 hours with approximately 5 ng/ml TNF alpha and 1ng/ml IL-1 beta, while NCI-H292 cells were activated for 6 and 14 hours with the following
5 cytokines: 5ng/ml IL-4, 5ng/ml IL-9, 5ng/ml IL-13 and 25ng/ml IFN gamma.

For these cell lines and blood cells, RNA was prepared by lysing approximately 10^7 cells/ml using Trizol (Gibco BRL). Briefly, 1/10 volume of bromochloropropane (Molecular Research Corporation) was added to the RNA sample, vortexed and after 10 minutes at room temperature, the tubes were spun at 14,000 rpm in a Sorvall SS34 rotor. The
10 aqueous phase was removed and placed in a 15ml Falcon Tube. An equal volume of isopropanol was added and left at -20 °C overnight. The precipitated RNA was spun down at 9,000 rpm for 15 min in a Sorvall SS34 rotor and washed in 70% ethanol. The pellet was redissolved in 300µl of RNase-free water and 35µl buffer (Promega) 5µl DTT, 7µl RNasin and 8µl DNase were added. The tube was incubated at 37 °C for 30 minutes to remove
15 contaminating genomic DNA, extracted once with phenol chloroform and re-precipitated with 1/10 volume of 3M sodium acetate and 2 volumes of 100% ethanol. The RNA was spun down and placed in RNase free water. RNA was stored at -80 °C.

AI_comprehensive panel_v1.0

The plates for AI_comprehensive panel_v1.0 include two control wells and 89 test
20 samples comprised of cDNA isolated from surgical and postmortem human tissues obtained from the Backus Hospital and Clinomics (Frederick, MD). Total RNA was extracted from tissue samples from the Backus Hospital in the Facility at CuraGen. Total RNA from other tissues was obtained from Clinomics.

Joint tissues including synovial fluid, synovium, bone and cartilage were obtained
25 from patients undergoing total knee or hip replacement surgery at the Backus Hospital. Tissue samples were immediately snap frozen in liquid nitrogen to ensure that isolated RNA was of optimal quality and not degraded. Additional samples of osteoarthritis and rheumatoid arthritis joint tissues were obtained from Clinomics. Normal control tissues were supplied by Clinomics and were obtained during autopsy of trauma victims.

30 Surgical specimens of psoriatic tissues and adjacent matched tissues were provided as total RNA by Clinomics. Two male and two female patients were selected between the ages

of 25 and 47. None of the patients were taking prescription drugs at the time samples were isolated.

Surgical specimens of diseased colon from patients with ulcerative colitis and Crohns disease and adjacent matched tissues were obtained from Clinomics. Bowel tissue from three female and three male Crohn's patients between the ages of 41-69 were used. Two patients were not on prescription medication while the others were taking dexamethasone, phenobarbital, or tylenol. Ulcerative colitis tissue was from three male and four female patients. Four of the patients were taking lebid and two were on phenobarbital.

Total RNA from post mortem lung tissue from trauma victims with no disease or with emphysema, asthma or COPD was purchased from Clinomics. Emphysema patients ranged in age from 40-70 and all were smokers, this age range was chosen to focus on patients with cigarette-linked emphysema and to avoid those patients with alpha-1 anti-trypsin deficiencies. Asthma patients ranged in age from 36-75, and excluded smokers to prevent those patients that could also have COPD. COPD patients ranged in age from 35-80 and included both smokers and non-smokers. Most patients were taking corticosteroids, and bronchodilators.

In the labels employed to identify tissues in the AI_comprehensive panel_v1.0 panel, the following abbreviations are used:

AI = Autoimmunity

Syn = Synovial

Normal = No apparent disease

Rep22 /Rep20 = individual patients

RA = Rheumatoid arthritis

Backus = From Backus Hospital

OA = Osteoarthritis

(SS) (BA) (MF) = Individual patients

Adj = Adjacent tissue

Match control = adjacent tissues

-M = Male

-F = Female

COPD = Chronic obstructive pulmonary disease

Panels 5D and 5I

The plates for Panel 5D and 5I include two control wells and a variety of cDNAs isolated from human tissues and cell lines with an emphasis on metabolic diseases. Metabolic tissues were obtained from patients enrolled in the Gestational Diabetes study. Cells were
 5 obtained during different stages in the differentiation of adipocytes from human mesenchymal stem cells. Human pancreatic islets were also obtained.

In the Gestational Diabetes study subjects are young (18 - 40 years), otherwise healthy women with and without gestational diabetes undergoing routine (elective) Caesarean section. After delivery of the infant, when the surgical incisions were being repaired/closed,
 10 the obstetrician removed a small sample (<1 cc) of the exposed metabolic tissues during the closure of each surgical level. The biopsy material was rinsed in sterile saline, blotted and fast frozen within 5 minutes from the time of removal. The tissue was then flash frozen in liquid nitrogen and stored, individually, in sterile screw-top tubes and kept on dry ice for shipment to or to be picked up by CuraGen. The metabolic tissues of interest include uterine
 15 wall (smooth muscle), visceral adipose, skeletal muscle (rectus) and subcutaneous adipose. Patient descriptions are as follows:

	Patient 2	Diabetic Hispanic, overweight, not on insulin
	Patient 7-9	Nondiabetic Caucasian and obese (BMI>30)
	Patient 10	Diabetic Hispanic, overweight, on insulin
20	Patient 11	Nondiabetic African American and overweight
	Patient 12	Diabetic Hispanic on insulin

Adipocyte differentiation was induced in donor progenitor cells obtained from Osirus (a division of Clonetics/BioWhittaker) in triplicate, except for Donor 3U which had only two replicates. Scientists at Clonetics isolated, grew and differentiated human mesenchymal stem
 25 cells (HuMSCs) for CuraGen based on the published protocol found in Mark F. Pittenger, et al., Multilineage Potential of Adult Human Mesenchymal Stem Cells Science Apr 2 1999: 143-147. Clonetics provided Trizol lysates or frozen pellets suitable for mRNA isolation and ds cDNA production. A general description of each donor is as follows:

	Donor 2 and 3 U: Mesenchymal Stem cells, Undifferentiated Adipose
30	Donor 2 and 3 AM: Adipose, AdiposeMidway Differentiated
	Donor 2 and 3 AD: Adipose, Adipose Differentiated

Human cell lines were generally obtained from ATCC (American Type Culture Collection), NCI or the German tumor cell bank and fall into the following tissue groups:

kidney proximal convoluted tubule, uterine smooth muscle cells, small intestine, liver HepG2 cancer cells, heart primary stromal cells, and adrenal cortical adenoma cells. These cells are all cultured under standard recommended conditions and RNA extracted using the standard procedures. All samples were processed at CuraGen to produce single stranded cDNA.

5 Panel 5I contains all samples previously described with the addition of pancreatic islets from a 58 year old female patient obtained from the Diabetes Research Institute at the University of Miami School of Medicine. Islet tissue was processed to total RNA at an outside source and delivered to CuraGen for addition to panel 5I.

10 In the labels employed to identify tissues in the 5D and 5I panels, the following abbreviations are used:

GO Adipose = Greater Omentum Adipose

SK = Skeletal Muscle

UT = Uterus

PL = Placenta

15 AD = Adipose Differentiated

AM = Adipose Midway Differentiated

U = Undifferentiated Stem Cells

Panel CNSD.01

20 The plates for Panel CNSD.01 include two control wells and 94 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center. Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

25 Disease diagnoses are taken from patient records. The panel contains two brains from each of the following diagnoses: Alzheimer's disease, Parkinson's disease, Huntington's disease, Progressive Supranuclear Palsy, Depression, and "Normal controls". Within each of these brains, the following regions are represented: cingulate gyrus, temporal pole, globus pallidus, substantia nigra, Brodman Area 4 (primary motor strip), Brodman Area 7 (parietal
30 cortex), Brodman Area 9 (prefrontal cortex), and Brodman area 17 (occipital cortex). Not all brain regions are represented in all cases; e.g., Huntington's disease is characterized in part by neurodegeneration in the globus pallidus, thus this region is impossible to obtain from

confirmed Huntington's cases. Likewise Parkinson's disease is characterized by degeneration of the substantia nigra making this region more difficult to obtain. Normal control brains were examined for neuropathology and found to be free of any pathology consistent with neurodegeneration.

- 5 In the labels employed to identify tissues in the CNS panel, the following abbreviations are used:

PSP = Progressive supranuclear palsy

Sub Nigra = Substantia nigra

Glob Palladus= Globus palladus

10 Temp Pole = Temporal pole

Cing Gyr = Cingulate gyrus

BA 4 = Brodman Area 4

Panel CNS_Neurodegeneration_V1.0

The plates for Panel CNS_Neurodegeneration_V1.0 include two control wells and 47
15 test samples comprised of cDNA isolated from postmortem human brain tissue obtained from the Harvard Brain Tissue Resource Center (McLean Hospital) and the Human Brain and Spinal Fluid Resource Center (VA Greater Los Angeles Healthcare System). Brains are removed from calvaria of donors between 4 and 24 hours after death, sectioned by neuroanatomists, and frozen at -80°C in liquid nitrogen vapor. All brains are sectioned and
20 examined by neuropathologists to confirm diagnoses with clear associated neuropathology.

Disease diagnoses are taken from patient records. The panel contains six brains from Alzheimer's disease (AD) patients, and eight brains from "Normal controls" who showed no evidence of dementia prior to death. The eight normal control brains are divided into two categories: Controls with no dementia and no Alzheimer's like pathology (Controls) and
25 controls with no dementia but evidence of severe Alzheimer's like pathology, (specifically senile plaque load rated as level 3 on a scale of 0-3; 0 = no evidence of plaques, 3 = severe AD senile plaque load). Within each of these brains, the following regions are represented: hippocampus, temporal cortex (Brodman Area 21), parietal cortex (Brodman area 7), and occipital cortex (Brodman area 17). These regions were chosen to encompass all levels of
30 neurodegeneration in AD. The hippocampus is a region of early and severe neuronal loss in AD; the temporal cortex is known to show neurodegeneration in AD after the hippocampus; the parietal cortex shows moderate neuronal death in the late stages of the disease; the

occipital cortex is spared in AD and therefore acts as a "control" region within AD patients.
Not all brain regions are represented in all cases.

In the labels employed to identify tissues in the CNS_Neurodegeneration_V1.0 panel, the following abbreviations are used:

- 5 AD = Alzheimer's disease brain; patient was demented and showed AD-like pathology upon autopsy
Control = Control brains; patient not demented, showing no neuropathology
Control (Path) = Control brains; patient not demented but showing severe AD-like pathology
10 SupTemporal Ctx = Superior Temporal Cortex
Inf Temporal Ctx = Inferior Temporal Cortex

A. NOV7a: Hsapiens CAB3

- Expression of gene NOV7a was assessed using the primer-probe set Ag4264, described in Table AA. Results of the RTQ-PCR runs are shown in Tables AB and AC.

Table AA. Probe Name Ag4264

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gcatcctgtcattctgttctt-3'	22	2294	345
Probe	TET-5'-tccctcatacatcttggagaaccgg-3'-TAMRA	26	2317	346
Reverse	5'-cagctatgagtcagggacaaca-3'	22	2352	347

Table AB. General screening panel v1.4

Tissue Name	Rel. Exp.(%) Ag4264, Run 222171301	Tissue Name	Rel. Exp.(%) Ag4264, Run 222171301
Adipose	4.0	Renal ca. TK-10	59.9
Melanoma* Hs688(A).T	24.7	Bladder	7.7
Melanoma* Hs688(B).T	18.7	Gastric ca. (liver met.) NCI-N87	49.3
Melanoma* M14	31.2	Gastric ca. KATO III	36.9
Melanoma* LOXIMVI	22.7	Colon ca. SW-948	15.1
Melanoma* SK-MEL-5	23.7	Colon ca. SW480	79.0
Squamous cell carcinoma SCC-4	8.1	Colon ca.* (SW480 met) SW620	10.2
Testis Pool	3.1	Colon ca. HT29	31.4
Prostate ca.* (bone met) PC-3	34.2	Colon ca. HCT-116	41.5
Prostate Pool	4.6	Colon ca. CaCo-2	25.7

Placenta	4.1	Colon cancer tissue	31.6
Uterus Pool	8.7	Colon ca. SW1116	12.3
Ovarian ca. OVCAR-3	79.0	Colon ca. Colo-205	2.8
Ovarian ca. SK-OV-3	40.3	Colon ca. SW-48	14.3
Ovarian ca. OVCAR-4	13.6	Colon Pool	55.9
Ovarian ca. OVCAR-5	43.2	Small Intestine Pool	27.4
Ovarian ca. IGROV-1	33.7	Stomach Pool	22.4
Ovarian ca. OVCAR-8	49.7	Bone Marrow Pool	15.6
Ovary	19.9	Fetal Heart	1.5
Breast ca. MCF-7	48.6	Heart Pool	9.9
Breast ca. MDA-MB-231	8.1	Lymph Node Pool	49.7
Breast ca. BT 549	85.3	Fetal Skeletal Muscle	2.1
Breast ca. T47D	100.0	Skeletal Muscle Pool	1.1
Breast ca. MDA-N	11.7	Spleen Pool	1.9
Breast Pool	35.6	Thymus Pool	19.9
Trachea	4.7	CNS cancer (glio/astro) U87-MG	19.2
Lung	33.9	CNS cancer (glio/astro) U-118-MG	59.0
Fetal Lung	11.0	CNS cancer (neuro;met) SK-N-AS	85.9
Lung ca. NCI-N417	2.7	CNS cancer (astro) SF-539	15.9
Lung ca. LX-1	52.5	CNS cancer (astro) SNB-75	35.6
Lung ca. NCI-H146	34.6	CNS cancer (glio) SNB-19	38.7
Lung ca. SHP-77	20.9	CNS cancer (glio) SF-295	67.8
Lung ca. A549	31.0	Brain (Amygdala) Pool	27.9
Lung ca. NCI-H526	8.1	Brain (cerebellum)	47.0
Lung ca. NCI-H23	50.0	Brain (fetal)	63.3
Lung ca. NCI-H460	12.1	Brain (Hippocampus) Pool	22.5
Lung ca. HOP-62	26.4	Cerebral Cortex Pool	42.9
Lung ca. NCI-H522	92.0	Brain (Substantia nigra) Pool	43.5
Liver	0.2	Brain (Thalamus) Pool	38.7
Fetal Liver	0.7	Brain (whole)	35.1
Liver ca. HepG2	0.5	Spinal Cord Pool	3.5
Kidney Pool	53.2	Adrenal Gland	1.4
Fetal Kidney	8.5	Pituitary gland Pool	2.3
Renal ca. 786-0	66.9	Salivary Gland	1.4
Renal ca. A498	12.0	Thyroid (female)	5.5
Renal ca. ACHN	56.3	Pancreatic ca. CAPAN2	29.1
Renal ca. UO-31	70.7	Pancreas Pool	42.3

Table AC. Panel 5 Islet

Tissue Name	Rel. Exp.(%) Ag4264, Run 181325881	Tissue Name	Rel. Exp.(%) Ag4264, Run 181325881
97457_Patient-02go_adipose	14.7	94709_Donor 2 AM - A_adipose	32.1
97476_Patient-07sk_skeletal muscle	13.3	94710_Donor 2 AM - B_adipose	14.1
97477_Patient-07ut_uterus	26.4	94711_Donor 2 AM - C_adipose	7.0
97478_Patient-07pl_placenta	8.6	94712_Donor 2 AD - A_adipose	18.4
99167_Bayer Patient 1	100.0	94713_Donor 2 AD - B_adipose	25.5
97482_Patient-08ut_uterus	6.0	94714_Donor 2 AD - C_adipose	24.5
97483_Patient-08pl_placenta	5.3	94742_Donor 3 U - A_Mesenchymal Stem Cells	5.6
97486_Patient-09sk_skeletal muscle	0.0	94743_Donor 3 U - B_Mesenchymal Stem Cells	12.4
97487_Patient-09ut_uterus	28.5	94730_Donor 3 AM - A_adipose	22.5
97488_Patient-09pl_placenta	7.2	94731_Donor 3 AM - B_adipose	12.0
97492_Patient-10ut_uterus	27.7	94732_Donor 3 AM - C_adipose	8.0
97493_Patient-10pl_placenta	21.0	94733_Donor 3 AD - A_adipose	23.3
97495_Patient-11go_adipose	7.4	94734_Donor 3 AD - B_adipose	2.4
97496_Patient-11sk_skeletal muscle	3.0	94735_Donor 3 AD - C_adipose	12.2
97497_Patient-11ut_uterus	24.0	77138_Liver_HepG2untreated	0.0
97498_Patient-11pl_placenta	15.2	73556_Heart_Cardiac stromal cells (primary)	19.9
97500_Patient-12go_adipose	15.6	81735_Small Intestine	11.0
97501_Patient-12sk_skeletal muscle	13.3	72409_Kidney_Proximal Convoluted Tubule	13.4
97502_Patient-12ut_uterus	36.6	82685_Small intestine_Duodenum	6.0
97503_Patient-12pl_placenta	7.4	90650_Adrenal_Adrenocortical adenoma	1.3
94721_Donor 2 U - A_Mesenchymal Stem Cells	28.5	72410_Kidney_HRCE	55.5
94722_Donor 2 U - B_Mesenchymal Stem Cells	18.9	72411_Kidney_HRE	52.5
94723_Donor 2 U - C_Mesenchymal Stem Cells	43.2	73139_Uterus_Uterine smooth muscle cells	27.4

General_screening_panel_v1.4 Summary: Ag4264 Highest expression of this gene is detected in breast cancer T47D cell line (CT=27). High to moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene

may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 5 Islet Summary: Ag4264 Highest expression of this gene is detected in islet cells (CT=32.3). Moderate to low levels of expression of this gene is also seen in uterus, mesenchymal stem cells, adipose and kidney.

This gene codes for the L-type calcium channel beta-3 subunit. The beta subunit of voltage-dependent calcium channels contributes to the function of the calcium channel by increasing peak calcium current, shifting the voltage dependencies of activation and inactivation, modulating G protein inhibition and controlling the alpha-1 subunit membrane targeting. Therefore, therapeutic modulation of this gene may be useful as a treatment for the enhancement of insulin secretion in Type 2 diabetes.

B. NOV9b: BHLH PROTEIN DEC2

Expression of full-length physical clone NOV9b was assessed using the primer-probe set Ag6927, described in Table BA. Results of the RTQ-PCR runs are shown in Table BB.

Table BA. Probe Name Ag6927

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-accgattctctgcccttc-3'	18	517	348
Probe	TET-5'-ctctgctcaagaagatccctcgagc-3'-TAMRA	26	567	349
Reverse	5'-gcaaggattcaggagcctt-3'	19	602	350

Table BB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6927, Run 278700378	Tissue Name	Rel. Exp.(%) Ag6927, Run 278700378
Adipose	4.5	Renal ca. TK-10	2.5
Melanoma* Hs688(A).T	8.5	Bladder	18.4
Melanoma* Hs688(B).T	5.5	Gastric ca. (liver met.) NCI-N87	23.3
Melanoma* M14	0.0	Gastric ca. KATO III	30.4
Melanoma* LOXIMVI	2.1	Colon ca. SW-948	11.4
Melanoma* SK-MEL-5	22.5	Colon ca. SW480	2.4
Squamous cell carcinoma SCC-4	10.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	6.6	Colon ca. HT29	41.2
Prostate ca.* (bone met) PC-3	22.5	Colon ca. HCT-116	4.4
Prostate Pool	2.5	Colon ca. CaCo-2	0.0
Placenta	4.8	Colon cancer tissue	14.9
Uterus Pool	1.7	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	100.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	5.1	Colon ca. SW-48	2.5
Ovarian ca. OVCAR-4	0.0	Colon Pool	9.9
Ovarian ca. OVCAR-5	46.7	Small Intestine Pool	9.0
Ovarian ca. IGROV-1	9.5	Stomach Pool	3.9
Ovarian ca. OVCAR-8	14.6	Bone Marrow Pool	8.1
Ovary	3.1	Fetal Heart	2.1
Breast ca. MCF-7	0.0	Heart Pool	1.4
Breast ca. MDA-MB-231	12.4	Lymph Node Pool	2.3
Breast ca. BT 549	1.9	Fetal Skeletal Muscle	4.7
Breast ca. T47D	0.0	Skeletal Muscle Pool	15.6
Breast ca. MDA-N	0.0	Spleen Pool	11.0
Breast Pool	3.9	Thymus Pool	7.0
Trachea	3.1	CNS cancer (glio/astro) U87-MG	43.2
Lung	21.3	CNS cancer (glio/astro) U-118-MG	13.5
Fetal Lung	3.6	CNS cancer (neuro;met) SK-N-AS	1.6
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	10.2
Lung ca. LX-1	1.1	CNS cancer (astro) SNB-75	37.9
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	13.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	57.4
Lung ca. A549	0.0	Brain (Amygdala) Pool	36.6
Lung ca. NCI-H526	3.1	Brain (cerebellum)	66.4
Lung ca. NCI-H23	6.3	Brain (fetal)	1.4
Lung ca. NCI-H460	1.6	Brain (Hippocampus) Pool	45.4
Lung ca. HOP-62	46.3	Cerebral Cortex Pool	40.6
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	26.2

Liver	0.0	Brain (Thalamus) Pool	60.7
Fetal Liver	0.0	Brain (whole)	14.5
Liver ca. HepG2	1.3	Spinal Cord Pool	89.5
Kidney Pool	21.9	Adrenal Gland	15.8
Fetal Kidney	1.4	Pituitary gland Pool	25.2
Renal ca. 786-0	6.3	Salivary Gland	9.3
Renal ca. A498	0.0	Thyroid (female)	9.9
Renal ca. ACHN	12.9	Pancreatic ca. CAPAN2	10.1
Renal ca. UO-31	80.7	Pancreas Pool	14.7

General_screening_panel_v1.6 Summary: Ag6927 Highest expression of this gene is detected in a ovarian cancer OVCAR-3 cell line (CT=33). Significant expression of this gene is also seen in number of cell lines derived from brain, colon, renal, lung and ovarian cancers. Therefore, expression of this gene may be used as marker to detect these cancers. Furthermore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

In addition, low levels of expression of this gene is also seen in most of the regions of the central nervous system examined, including amygdala, hippocampus, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

C. NOV16a: FORKHEAD PROTEIN O3A

Expression of gene NOV16a was assessed using the primer-probe set Ag3742, described in Table CA. Results of the RTQ-PCR runs are shown in Table CB.

Table CA. Probe Name Ag3742

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agaccctcaaactgacacaaga-3'	22	1863	351
Probe	TET-5'-aaaacccttgccaaatctgcttca-3'-TAMRA	26	1894	352
Reverse	5'-aacgggtatcactgtccacttg-3'	22	1921	353

Table CB. Panel 5D

Tissue Name	Rel. Exp.(%) Ag3742, Run 169315028	Tissue Name	Rel. Exp.(%) Ag3742, Run 169315028
97457_Patient-02go_adipose	57.8	94709_Donor 2 AM - A_adipose	44.1
97476_Patient-07sk_skeletal muscle	29.9	94710_Donor 2 AM - B_adipose	29.9
97477_Patient-07ut_uterus	20.0	94711_Donor 2 AM - C_adipose	22.1
97478_Patient-07pl_placenta	32.3	94712_Donor 2 AD - A_adipose	41.8
97481_Patient-08sk_skeletal muscle	41.5	94713_Donor 2 AD - B_adipose	63.7
97482_Patient-08ut_uterus	14.3	94714_Donor 2 AD - C_adipose	55.5
97483_Patient-08pl_placenta	37.1	94742_Donor 3 U - A_Mesenchymal Stem Cells	19.9
97486_Patient-09sk_skeletal muscle	17.2	94743_Donor 3 U - B_Mesenchymal Stem Cells	21.0
97487_Patient-09ut_uterus	38.2	94730_Donor 3 AM - A_adipose	52.1
97488_Patient-09pl_placenta	13.5	94731_Donor 3 AM - B_adipose	32.3
97492_Patient-10ut_uterus	36.6	94732_Donor 3 AM - C_adipose	33.2
97493_Patient-10pl_placenta	43.2	94733_Donor 3 AD - A_adipose	41.8
97495_Patient-11go_adipose	36.6	94734_Donor 3 AD - B_adipose	19.9
97496_Patient-11sk_skeletal muscle	66.4	94735_Donor 3 AD - C_adipose	44.1
97497_Patient-11ut_uterus	40.3	77138_Liver_HepG2untreated	100.0
97498_Patient-11pl_placenta	17.7	73556_Heart_Cardiac stromal cells (primary)	14.1
97500_Patient-12go_adipose	44.8	81735_Small Intestine	47.6
97501_Patient-12sk_skeletal muscle	77.4	72409_Kidney_Proximal Convoluted Tubule	17.8
97502_Patient-12ut_uterus	30.1	82685_Small intestine_Duodenum	20.9
97503_Patient-12pl_placenta	12.2	90650_Adrenal Adrenocortical adenoma	6.6
94721_Donor 2 U - A_Mesenchymal Stem Cells	23.8	72410_Kidney_HRCE	43.5
94722_Donor 2 U - B_Mesenchymal Stem Cells	19.5	72411_Kidney_HRE	49.0
94723_Donor 2 U - C_Mesenchymal Stem Cells	20.2	73139_Uterus_Uterine smooth muscle cells	30.6

Panel 5D Summary: Ag3742 Highest expression of this gene is detected in liver HepG2 cell line (CT=28.8). This gene shows a wide expression in tissues with metabolic/endocrine function. Moderate to low levels of expression of this gene is seen in

5 adipose, uterus, placenta, skeletal muscle, small intestine and kidney. Therefore, therapeutic

modulation of this gene may be useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

D. NOV18a: kinectin like

Expression of gene NOV18a was assessed using the primer-probe set Ag6564,
5 described in Table DA. Results of the RTQ-PCR runs are shown in Tables DB and DC.

Table DA. Probe Name Ag6564

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttgcacctgttcattgaat-3'	20	359	354
Probe	TET-5'-tcctaactactctgaagttcaacga-3'-TAMRA	28	380	355
Reverse	5'-tcttcaagcacaggcttttg-3'	20	430	356

Table DB. AI comprehensive panel v1.0

10

Tissue Name	Rel. Exp.(%) Ag6564, Run 277314696	Tissue Name	Rel. Exp.(%) Ag6564, Run 277314696
110967 COPD-F	19.2	112427 Match Control Psoriasis-F	100.0
110980 COPD-F	20.4	112418 Psoriasis-M	5.2
110968 COPD-M	27.0	112723 Match Control Psoriasis-M	13.1
110977 COPD-M	58.2	112419 Psoriasis-M	14.0
110989 Emphysema-F	44.1	112424 Match Control Psoriasis-M	16.2
110992 Emphysema-F	22.5	112420 Psoriasis-M	65.1
110993 Emphysema-F	29.7	112425 Match Control Psoriasis-M	46.7
110994 Emphysema-F	14.3	104689 (MF) OA Bone-Backus	33.7
110995 Emphysema-F	40.1	104690 (MF) Adj "Normal" Bone-Backus	33.0
110996 Emphysema-F	8.0	104691 (MF) OA Synovium-Backus	30.6
110997 Asthma-M	4.1	104692 (BA) OA Cartilage-Backus	13.9
111001 Asthma-F	32.1	104694 (BA) OA Bone-Backus	24.7
111002 Asthma-F	42.3	104695 (BA) Adj "Normal" Bone-Backus	32.8
111003 Atopic Asthma-F	38.2	104696 (BA) OA Synovium-Backus	26.6
111004 Atopic Asthma-F	35.4	104700 (SS) OA Bone-Backus	26.2
111005 Atopic Asthma-F	24.7	104701 (SS) Adj "Normal" Bone-Backus	23.5
111006 Atopic Asthma-F	6.4	104702 (SS) OA Synovium-Backus	46.0
111417 Allergy-M	24.0	117093 OA Cartilage Rep7	34.4
112347 Allergy-M	0.0	112672 OA Bone5	24.5
112349 Normal Lung-F	0.0	112673 OA Synovium5	14.5

112357 Normal Lung-F	48.0	112674 OA Synovial Fluid cells5	20.2
112354 Normal Lung-M	21.9	117100 OA Cartilage Rep14	4.3
112374 Crohns-F	31.6	112756 OA Bone9	74.7
112389 Match Control Crohns-F	22.8	112757 OA Synovium9	33.4
112375 Crohns-F	24.1	112758 OA Synovial Fluid Cells9	17.4
112732 Match Control Crohns-F	13.4	117125 RA Cartilage Rep2	5.1
112725 Crohns-M	5.5	113492 Bone2 RA	14.6
112387 Match Control Crohns-M	7.2	113493 Synovium2 RA	5.7
112378 Crohns-M	0.0	113494 Syn Fluid Cells RA	11.5
112390 Match Control Crohns-M	55.9	113499 Cartilage4 RA	10.8
112726 Crohns-M	26.6	113500 Bone4 RA	13.8
112731 Match Control Crohns-M	21.6	113501 Synovium4 RA	9.2
112380 Ulcer Col-F	39.8	113502 Syn Fluid Cells4 RA	9.9
112734 Match Control Ulcer Col-F	42.3	113495 Cartilage3 RA	9.2
112384 Ulcer Col-F	49.7	113496 Bone3 RA	10.6
112737 Match Control Ulcer Col-F	3.2	113497 Synovium3 RA	8.2
112386 Ulcer Col-F	19.9	113498 Syn Fluid Cells3 RA	13.2
112738 Match Control Ulcer Col-F	8.2	117106 Normal Cartilage Rep20	1.3
112381 Ulcer Col-M	0.0	113663 Bone3 Normal	0.0
112735 Match Control Ulcer Col-M	1.1	113664 Synovium3 Normal	0.0
112382 Ulcer Col-M	21.3	113665 Syn Fluid Cells3 Normal	0.0
112394 Match Control Ulcer Col-M	7.2	117107 Normal Cartilage Rep22	5.3
112383 Ulcer Col-M	34.9	113667 Bone4 Normal	17.3
112736 Match Control Ulcer Col-M	3.9	113668 Synovium4 Normal	18.9
112423 Psoriasis-F	17.8	113669 Syn Fluid Cells4 Normal	25.3

Table DC. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6564, Run 277243357	Tissue Name	Rel. Exp.(%) Ag6564, Run 277243357
Adipose	19.6	Renal ca. TK-10	31.9
Melanoma* Hs688(A).T	19.9	Bladder	30.8
Melanoma* Hs688(B).T	21.5	Gastric ca. (liver met.) NCI-N87	53.2
Melanoma* M14	8.4	Gastric ca. KATO III	39.2
Melanoma* LOXIMVI	15.0	Colon ca. SW-948	6.2
Melanoma* SK-MEL-5	35.1	Colon ca. SW480	47.0

Squamous cell carcinoma SCC-4	39.0	Colon ca.* (SW480 met) SW620	22.1
Testis Pool	25.0	Colon ca. HT29	11.8
Prostate ca.* (bone met) PC-3	87.1	Colon ca. HCT-116	26.8
Prostate Pool	17.8	Colon ca. CaCo-2	28.9
Placenta	1.5	Colon cancer tissue	20.0
Uterus Pool	7.5	Colon ca. SW1116	6.6
Ovarian ca. OVCAR-3	48.3	Colon ca. Colo-205	2.6
Ovarian ca. SK-OV-3	100.0	Colon ca. SW-48	3.0
Ovarian ca. OVCAR-4	9.9	Colon Pool	24.5
Ovarian ca. OVCAR-5	11.3	Small Intestine Pool	21.6
Ovarian ca. IGROV-1	15.9	Stomach Pool	11.0
Ovarian ca. OVCAR-8	13.7	Bone Marrow Pool	9.9
Ovary	7.2	Fetal Heart	34.2
Breast ca. MCF-7	32.8	Heart Pool	17.4
Breast ca. MDA-MB-231	67.8	Lymph Node Pool	29.3
Breast ca. BT 549	94.6	Fetal Skeletal Muscle	12.9
Breast ca. T47D	17.8	Skeletal Muscle Pool	24.8
Breast ca. MDA-N	7.9	Spleen Pool	15.4
Breast Pool	26.8	Thymus Pool	19.9
Trachea	10.3	CNS cancer (glio/astro) U87-MG	71.2
Lung	9.1	CNS cancer (glio/astro) U-118-MG	46.0
Fetal Lung	49.7	CNS cancer (neuro;met) SK-N-AS	53.6
Lung ca. NCI-N417	4.2	CNS cancer (astro) SF-539	18.2
Lung ca. LX-1	26.2	CNS cancer (astro) SNB-75	62.9
Lung ca. NCI-H146	5.4	CNS cancer (glio) SNB-19	16.7
Lung ca. SHP-77	46.3	CNS cancer (glio) SF-295	34.4
Lung ca. A549	27.2	Brain (Amygdala) Pool	8.7
Lung ca. NCI-H526	7.7	Brain (cerebellum)	19.1
Lung ca. NCI-H23	40.3	Brain (fetal)	8.0
Lung ca. NCI-H460	20.6	Brain (Hippocampus) Pool	12.4
Lung ca. HOP-62	13.6	Cerebral Cortex Pool	6.1
Lung ca. NCI-H522	17.2	Brain (Substantia nigra) Pool	6.0
Liver	0.0	Brain (Thalamus) Pool	17.4
Fetal Liver	15.9	Brain (whole)	4.9
Liver ca. HepG2	14.6	Spinal Cord Pool	16.6
Kidney Pool	25.9	Adrenal Gland	7.4
Fetal Kidney	43.5	Pituitary gland Pool	15.8
Renal ca. 786-0	31.2	Salivary Gland	1.7
Renal ca. A498	18.3	Thyroid (female)	5.8
Renal ca. ACHN	10.7	Pancreatic ca. CAPAN2	35.1
Renal ca. UO-31	26.1	Pancreas Pool	17.0

AI_comprehensive_panel_v1.0 Summary: Ag6564 Highest expression of this gene is detected in control psoriasis sample (CT=24). This gene shows wide expression in this panel with high levels of expression of in samples derived from bone, cartilage, synovium and synovial fluid samples from osteoarthritis, and rheumatoid arthritis patients, as well as, in
5 samples derived from normal lung samples, COPD lung, emphysema, atopic asthma, asthma, allergy, Crohn's disease (normal matched control and diseased), ulcerative colitis (normal matched control and diseased), and psoriasis (normal matched control and diseased). Therefore, therapeutic modulation of this gene product may ameliorate symptoms/conditions associated with autoimmune and inflammatory disorders including psoriasis, allergy, asthma,
10 inflammatory bowel disease, rheumatoid arthritis and osteoarthritis

General_screening_panel_v1.6 Summary: Ag6564 Highest expression of this gene is seen in ovarian cancer SK-OV-3 cell line (CT=24.4). High levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus,
15 expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high
20 levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CT=27) when
25 compared to adult liver (CT=40). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in
30 treatment of liver related diseases.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum,

cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

E. NOV20a: XIN

- 5 Expression of gene NOV20a was assessed using the primer-probe set Ag3459, described in Table EA. Results of the RTQ-PCR runs are shown in Tables EB, EC and ED.

Table EA. Probe Name Ag3459

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acacaactggctcaggacatag-3'	22	4705	357
Probe	TET-5'-ctgctccaccagaaaggtgtccaag-3'-TAMRA	25	4735	358
Reverse	5'-gtgatgtccttcttccagttt-3'	22	4763	359

10 **Table EB. General screening panel v1.4**

Tissue Name	Rel. Exp.(%) Ag3459, Run 217066335	Tissue Name	Rel. Exp.(%) Ag3459, Run 217066335
Adipose	1.8	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.3	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.1
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	19.9
Breast ca. MCF-7	0.0	Heart Pool	4.2
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0

Breast ca. BT 549	0.0	Fetal Skeletal Muscle	12.4
Breast ca. T47D	0.0	Skeletal Muscle Pool	100.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	0.2	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.1
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

Table EC. Panel 4D

Tissue Name	Rel. Exp.(%) Ag3459, Run 166417097	Tissue Name	Rel. Exp.(%) Ag3459, Run 166417097
Secondary Th1 act	0.2	HUVEC IL-1beta	0.0
Secondary Th2 act	7.9	HUVEC IFN gamma	0.0
Secondary Tr1 act	7.7	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.1	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	2.2	Microvascular Dermal EC none	0.0

Primary Tr1 act	0.6	Microsvascular Dermal EC TNFalpha + IL-1 beta	0.0
Primary Th1 rest	0.1	Bronchial epithelium TNFalpha + IL1 beta	0.0
Primary Th2 rest	0.0	Small airway epithelium none	0.0
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1 beta	0.0
CD45RA CD4 lymphocyte act	0.2	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.5	Coronary artery SMC TNFalpha + IL-1 beta	0.0
CD8 lymphocyte act	0.6	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	1.1	Astrocytes TNFalpha + IL-1 beta	2.4
Secondary CD8 lymphocyte act	7.3	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	63.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	2.3	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	0.3
LAK cells IL-2	0.1	Liver cirrhosis	1.1
LAK cells IL-2+IL-12	0.7	Lupus kidney	0.0
LAK cells IL-2+IFN gamma	2.2	NCI-H292 none	0.0
LAK cells IL-2+ IL-18	1.2	NCI-H292 IL-4	0.0
LAK cells PMA/ionomycin	100.0	NCI-H292 IL-9	0.0
NK Cells IL-2 rest	0.1	NCI-H292 IL-13	0.0
Two Way MLR 3 day	1.5	NCI-H292 IFN gamma	0.0
Two Way MLR 5 day	7.9	HPAEC none	0.0
Two Way MLR 7 day	3.8	HPAEC TNF alpha + IL-1 beta	0.1
PBMC rest	0.0	Lung fibroblast none	0.3
PBMC PWM	0.7	Lung fibroblast TNF alpha + IL-1 beta	0.3
PBMC PHA-L	0.3	Lung fibroblast IL-4	1.6
Ramos (B cell) none	0.0	Lung fibroblast IL-9	0.3
Ramos (B cell) ionomycin	0.0	Lung fibroblast IL-13	1.0
B lymphocytes PWM	1.7	Lung fibroblast IFN gamma	32.1
B lymphocytes CD40L and IL-4	0.2	Dermal fibroblast CCD1070 rest	0.3
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 TNF alpha	0.8
EOL-1 dbcAMP PMA/ionomycin	32.5	Dermal fibroblast CCD1070 IL-1 beta	0.2
Dendritic cells none	0.5	Dermal fibroblast IFN gamma	0.2
Dendritic cells LPS	1.2	Dermal fibroblast IL-4	0.0
Dendritic cells anti-CD40	0.1	IBD Colitis 2	0.4
Monocytes rest	0.0	IBD Crohn's	0.3
Monocytes LPS	0.0	Colon	2.1
Macrophages rest	0.2	Lung	0.6
Macrophages LPS	8.8	Thymus	0.0

HUVEC none	0.0	Kidney	0.3
HUVEC starved	0.0		

Table ED. general oncology screening panel v 2.4

Tissue Name	Rel. Exp.(%) Ag3459, Run 267242232	Tissue Name	Rel. Exp.(%) Ag3459, Run 267242232
Colon cancer 1	5.9	Bladder cancer NAT 2	0.0
Colon cancer NAT 1	0.0	Bladder cancer NAT 3	2.1
Colon cancer 2	5.3	Bladder cancer NAT 4	0.0
Colon cancer NAT 2	1.9	Prostate adenocarcinoma 1	7.5
Colon cancer 3	8.0	Prostate adenocarcinoma 2	0.0
Colon cancer NAT 3	0.0	Prostate adenocarcinoma 3	0.0
Colon malignant cancer 4	24.1	Prostate adenocarcinoma 4	7.4
Colon normal adjacent tissue 4	2.4	Prostate cancer NAT 5	100.0
Lung cancer 1	60.3	Prostate adenocarcinoma 6	5.1
Lung NAT 1	1.4	Prostate adenocarcinoma 7	0.0
Lung cancer 2	71.7	Prostate adenocarcinoma 8	0.0
Lung NAT 2	1.5	Prostate adenocarcinoma 9	0.0
Squamous cell carcinoma 3	4.4	Prostate cancer NAT 10	0.0
Lung NAT 3	0.0	Kidney cancer 1	0.0
metastatic melanoma 1	6.0	Kidney NAT 1	1.5
Melanoma 2	8.7	Kidney cancer 2	9.1
Melanoma 3	0.0	Kidney NAT 2	6.0
metastatic melanoma 4	3.9	Kidney cancer 3	6.2
metastatic melanoma 5	10.4	Kidney NAT 3	0.0
Bladder cancer 1	4.4	Kidney cancer 4	5.4
Bladder cancer NAT 1	0.0	Kidney NAT 4	0.0
Bladder cancer 2	0.0		

- 5 **General_screening_panel_v1.4 Summary:** Ag3459 Highest expression of this gene is detected in skeletal muscle (CT=23.7). Interestingly, expression of this gene is higher in adult as compared to fetal skeletal muscle (CT=26.7). Therefore, expression of this gene may be used to distinguish between the adult and fetal skeletal muscle.

- 10 Moderate to low levels of expression of this gene is also seen in tissues with metabolic functions including adipose, skeletal muscle, heart, and liver. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of metabolically related diseases, such as obesity and diabetes.

Panel 4D Summary: Ag3459 Highest expression of this gene is detected in PMA/ionomycin stimulated LAK cells (CT=24.9). Moderate to low levels of expression of this gene is seen in activated polarized T cells, memory and naive T cells, activated B cells, two way MLR, LAK, eosinophils, dendritic cells, lung and dermal fibroblasts, colon, lung, kidney, liver cirrhosis and lupus kidney. Interestingly, expression of this gene is upregulated in cytokine activated LAK cells, polarized T cells, PBMC, eosinophils, macrophage, basophils, keratinocytes and HPAEC endothelial cells. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

general oncology screening panel_v_2.4 Summary: Ag3459 Highest expression of this gene is detected in normal adjacent prostate sample (CT=32.7). Expression of this gene is higher in normal tissue as compared to prostate cancer (CTs=36-40). Therefore, expression of this gene may be used to distinguish the cancerous region from normal prostate.

In addition, moderate to low levels of expression of this gene is also seen in malignant colon cancer and lung cancers. Expression of this gene is higher in cancer as compared to the corresponding normal adjacent tissue. Therefore, expression of this gene may be used as marker to detect the presence of colon and lung cancer. In addition, therapeutic modulation of this gene may be useful in the treatment of these cancers.

F. NOV21a: PROSTATIC BINDING PROTEIN

Expression of full-length physical clone NOV21a was assessed using the primer-probe set Ag4464, described in Table FA. Results of the RTQ-PCR runs are shown in Tables FB, FC and FD.

Table FA. Probe Name Ag4464

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cttgattcagggaagctctaca-3'	22	177	360
Probe	TET-5'-ctcccagcaggaaggatcccaat-3'-TAMRA	24	223	361
Reverse	5'-aggaaatgatgccattctctgt-3'	22	247	362

Table FB. General screening panel v1.4

Tissue Name	Rel. Exp.(%) Ag4464, Run 222653242	Tissue Name	Rel. Exp.(%) Ag4464, Run 222653242
Adipose	9.0	Renal ca. TK-10	25.7
Melanoma* Hs688(A).T	21.0	Bladder	14.0
Melanoma* Hs688(B).T	18.4	Gastric ca. (liver met.) NCI-N87	16.3
Melanoma* M14	34.4	Gastric ca. KATO III	35.6
Melanoma* LOXIMVI	29.9	Colon ca. SW-948	18.6
Melanoma* SK-MEL-5	60.3	Colon ca. SW480	34.4
Squamous cell carcinoma SCC-4	7.6	Colon ca.* (SW480 met) SW620	23.5
Testis Pool	9.1	Colon ca. HT29	19.2
Prostate ca.* (bone met) PC-3	27.5	Colon ca. HCT-116	39.0
Prostate Pool	6.8	Colon ca. CaCo-2	15.2
Placenta	3.8	Colon cancer tissue	18.0
Uterus Pool	5.4	Colon ca. SW1116	12.2
Ovarian ca. OVCAR-3	7.3	Colon ca. Colo-205	14.1
Ovarian ca. SK-OV-3	31.9	Colon ca. SW-48	15.7
Ovarian ca. OVCAR-4	15.4	Colon Pool	13.1
Ovarian ca. OVCAR-5	25.2	Small Intestine Pool	7.5
Ovarian ca. IGROV-1	19.6	Stomach Pool	5.5
Ovarian ca. OVCAR-8	11.7	Bone Marrow Pool	4.7
Ovary	15.8	Fetal Heart	9.6
Breast ca. MCF-7	42.3	Heart Pool	11.5
Breast ca. MDA-MB-231	30.8	Lymph Node Pool	11.5
Breast ca. BT 549	27.9	Fetal Skeletal Muscle	3.9
Breast ca. T47D	59.5	Skeletal Muscle Pool	17.3
Breast ca. MDA-N	14.3	Spleen Pool	11.7
Breast Pool	10.6	Thymus Pool	6.5
Trachea	12.9	CNS cancer (glio/astro) U87-MG	28.1
Lung	3.9	CNS cancer (glio/astro) U-118-MG	28.9
Fetal Lung	12.5	CNS cancer (neuro;met) SK-N-AS	26.1
Lung ca. NCI-N417	13.3	CNS cancer (astro) SF-539	8.5
Lung ca. LX-1	11.0	CNS cancer (astro) SNB-75	32.1
Lung ca. NCI-H146	14.9	CNS cancer (glio) SNB-19	21.0
Lung ca. SHP-77	35.6	CNS cancer (glio) SF-295	21.0
Lung ca. A549	24.7	Brain (Amygdala) Pool	31.9
Lung ca. NCI-H526	18.6	Brain (cerebellum)	74.2
Lung ca. NCI-H23	15.8	Brain (fetal)	23.2
Lung ca. NCI-H460	7.7	Brain (Hippocampus) Pool	36.3
Lung ca. HOP-62	7.1	Cerebral Cortex Pool	58.6
Lung ca. NCI-H522	17.4	Brain (Substantia nigra) Pool	44.1
Liver	14.4	Brain (Thalamus) Pool	54.7

Fetal Liver	48.6	Brain (whole)	30.8
Liver ca. HepG2	39.8	Spinal Cord Pool	47.6
Kidney Pool	16.7	Adrenal Gland	100.0
Fetal Kidney	8.6	Pituitary gland Pool	7.4
Renal ca. 786-0	26.8	Salivary Gland	9.6
Renal ca. A498	5.0	Thyroid (female)	31.0
Renal ca. ACHN	14.7	Pancreatic ca. CAPAN2	6.7
Renal ca. UO-31	23.3	Pancreas Pool	14.7

Table FC. Panel CNS 1

Tissue Name	Rel. Exp.(%) Ag4464, Run 191785915	Tissue Name	Rel. Exp.(%) Ag4464, Run 191785915
BA4 Control	48.6	BA17 PSP	38.2
BA4 Control2	49.7	BA17 PSP2	16.0
BA4 Alzheimer's2	7.9	Sub Nigra Control	57.4
BA4 Parkinson's	54.0	Sub Nigra Control2	37.1
BA4 Parkinson's2	95.9	Sub Nigra Alzheimer's2	27.5
BA4 Huntington's	45.4	Sub Nigra Parkinson's2	93.3
BA4 Huntington's2	13.7	Sub Nigra Huntington's	94.6
BA4 PSP	18.0	Sub Nigra Huntington's2	51.1
BA4 PSP2	50.0	Sub Nigra PSP2	13.4
BA4 Depression	20.7	Sub Nigra Depression	12.2
BA4 Depression2	10.4	Sub Nigra Depression2	14.5
BA7 Control	66.0	Glob Palladus Control	16.2
BA7 Control2	45.1	Glob Palladus Control2	16.5
BA7 Alzheimer's2	12.0	Glob Palladus Alzheimer's	29.9
BA7 Parkinson's	22.1	Glob Palladus Alzheimer's2	12.6
BA7 Parkinson's2	47.0	Glob Palladus Parkinson's	84.7
BA7 Huntington's	60.7	Glob Palladus Parkinson's2	27.0
BA7 Huntington's2	39.2	Glob Palladus PSP	8.4
BA7 PSP	56.6	Glob Palladus PSP2	14.6
BA7 PSP2	55.9	Glob Palladus Depression	11.3
BA7 Depression	15.1	Temp Pole Control	16.7
BA9 Control	47.3	Temp Pole Control2	72.7
BA9 Control2	85.9	Temp Pole Alzheimer's	12.0
BA9 Alzheimer's	12.9	Temp Pole Alzheimer's2	8.4
BA9 Alzheimer's2	27.0	Temp Pole Parkinson's	39.0
BA9 Parkinson's	33.9	Temp Pole Parkinson's2	47.0
BA9 Parkinson's2	66.0	Temp Pole Huntington's	57.4
BA9 Huntington's	76.8	Temp Pole PSP	6.9

BA9 Huntington's2	28.1	Temp Pole PSP2	12.3
BA9 PSP	25.2	Temp Pole Depression2	8.5
BA9 PSP2	8.0	Cing Gyr Control	66.9
BA9 Depression	14.2	Cing Gyr Control2	40.3
BA9 Depression2	11.8	Cing Gyr Alzheimer's	44.4
BA17 Control	48.6	Cing Gyr Alzheimer's2	14.4
BA17 Control2	57.8	Cing Gyr Parkinson's	33.0
BA17 Alzheimer's2	8.2	Cing Gyr Parkinson's2	52.5
BA17 Parkinson's	38.7	Cing Gyr Huntington's	100.0
BA17 Parkinson's2	49.3	Cing Gyr Huntington's2	33.7
BA17 Huntington's	40.1	Cing Gyr PSP	36.6
BA17 Huntington's2	15.6	Cing Gyr PSP2	15.7
BA17 Depression	13.3	Cing Gyr Depression	10.2
BA17 Depression2	33.7	Cing Gyr Depression2	17.6

Table FD. Panel CNS 1.1

Tissue Name	Rel. Exp.(%) Ag4464, Run 195308648	Tissue Name	Rel. Exp.(%) Ag4464, Run 195308648
Cing Gyr Depression2	22.8	BA17 PSP2	24.8
Cing Gyr Depression	12.9	BA17 PSP	46.7
Cing Gyr PSP2	14.6	BA17 Huntington's2	14.5
Cing Gyr PSP	33.9	BA17 Huntington's	45.4
Cing Gyr Huntington's2	30.6	BA17 Parkinson's2	49.3
Cing Gyr Huntington's	97.9	BA17 Parkinson's	33.2
Cing Gyr Parkinson's2	66.9	BA17 Alzheimer's2	9.0
Cing Gyr Parkinson's	47.0	BA17 Control2	54.7
Cing Gyr Alzheimer's2	15.6	BA17 Control	46.3
Cing Gyr Alzheimer's	47.6	BA9 Depression2	14.5
Cing Gyr Control2	43.2	BA9 Depression	14.0
Cing Gyr Control	73.2	BA9 PSP2	10.8
Temp Pole Depression2	11.0	BA9 PSP	7.6
Temp Pole PSP2	14.1	BA9 Huntington's2	24.3
Temp Pole PSP	8.1	BA9 Huntington's	88.3
Temp Pole Huntington's	46.0	BA9 Parkinson's2	77.4
Temp Pole Parkinson's2	48.0	BA9 Parkinson's	45.1
Temp Pole Parkinson's	36.6	BA9 Alzheimer's2	24.0
Temp Pole Alzheimer's2	14.4	BA9 Alzheimer's	14.2
Temp Pole Alzheimer's	14.6	BA9 Control2	94.0
Temp Pole Control2	70.7	BA9 Control	48.0
Temp Pole Control	17.9	BA7 Depression	12.2

Glob Palladus Depression	11.0	BA7 PSP2	62.4
Glob Palladus PSP2	15.8	BA7 PSP	59.0
Glob Palladus PSP	10.4	BA7 Huntington's2	42.9
Glob Palladus Parkinson's2	28.9	BA7 Huntington's	55.1
Glob Palladus Parkinson's	88.9	BA7 Parkinson's2	33.7
Glob Palladus Alzheimer's2	13.6	BA7 Parkinson's	23.8
Glob Palladus Alzheimer's	37.4	BA7 Alzheimer's2	9.8
Glob Palladus Control2	13.3	BA7 Control2	42.9
Glob Palladus Control	17.0	BA7 Control	59.0
Sub Nigra Depression2	18.2	BA4 Depression2	12.7
Sub Nigra Depression	14.6	BA4 Depression	26.8
Sub Nigra PSP2	14.5	BA4 PSP2	53.2
Sub Nigra Huntington's2	50.3	BA4 PSP	22.4
Sub Nigra Huntington's	89.5	BA4 Huntington's2	15.2
Sub Nigra Parkinson's2	95.3	BA4 Huntington's	51.8
Sub Nigra Alzheimer's2	25.7	BA4 Parkinson's2	100.0
Sub Nigra Control2	39.0	BA4 Parkinson's	58.6
Sub Nigra Control	46.7	BA4 Alzheimer's2	11.2
BA17 Depression2	33.0	BA4 Control2	62.9
BA17 Depression	12.9	BA4 Control	49.7

General_screening_panel_v1.4 Summary: Ag4464 Highest expression of this gene is seen in adrenal gland (CT=23.4). In addition, this gene is also expressed at high levels in pancreas, adipose, thyroid, pituitary gland, skeletal muscle, heart, liver and the
5 gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum,
10 cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

High levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous
15 cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric,

colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Panel CNS_1 Summary: Ag4464 This panel confirms the expression of this gene at significant levels in the brains of an independent group of individuals. Please see Panel 1.4 for a discussion of the potential use of this gene in treatment of central nervous system disorders.

Panel CNS_1.1 Summary: Ag4464 This panel confirms the expression of this gene at significant levels in the brains of an independent group of individuals. Please see Panel 1.4 for a discussion of the potential use of this gene in treatment of central nervous system disorders.

G. NOV32b: EH DOMAIN-BINDING MITOTIC PHOSPHOPROTEIN

Expression of gene NOV32b was assessed using the primer-probe set Ag3088, described in Table GA. Results of the RTQ-PCR runs are shown in Tables GB, GC and GD.

Table GA. Probe Name Ag3088

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cacgtttacaaggccatgac-3'	20	1096	363
Probe	TET-5'-atggagtacatcaagaccggctc-3'-TAMRA	26	1120	364
Reverse	5'-atgttccttgactgctg-3'	20	1159	365

Table GB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag3088, Run 208974163	Tissue Name	Rel. Exp.(%) Ag3088, Run 208974163
AD 1 Hippo	19.8	Control (Path) 3 Temporal Ctx	14.0
AD 2 Hippo	35.6	Control (Path) 4 Temporal Ctx	44.4
AD 3 Hippo	17.9	AD 1 Occipital Ctx	27.9
AD 4 Hippo	17.4	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	15.6
AD 6 Hippo	62.0	AD 4 Occipital Ctx	92.7
Control 2 Hippo	58.2	AD 5 Occipital Ctx	71.2
Control 4 Hippo	15.3	AD 6 Occipital Ctx	26.1
Control (Path) 3 Hippo	13.5	Control 1 Occipital Ctx	8.7
AD 1 Temporal Ctx	29.1	Control 2 Occipital Ctx	77.9
AD 2 Temporal Ctx	47.6	Control 3 Occipital Ctx	29.7
AD 3 Temporal Ctx	14.8	Control 4 Occipital Ctx	9.8

AD 4 Temporal Ctx	34.2	Control (Path) 1 Occipital Ctx	69.3
AD 5 Inf Temporal Ctx	84.7	Control (Path) 2 Occipital Ctx	16.8
AD 5 Sup Temporal Ctx	47.6	Control (Path) 3 Occipital Ctx	7.1
AD 6 Inf Temporal Ctx	65.5	Control (Path) 4 Occipital Ctx	24.7
AD 6 Sup Temporal Ctx	60.3	Control 1 Parietal Ctx	13.9
Control 1 Temporal Ctx	10.0	Control 2 Parietal Ctx	66.9
Control 2 Temporal Ctx	69.3	Control 3 Parietal Ctx	19.9
Control 3 Temporal Ctx	33.9	Control (Path) 1 Parietal Ctx	63.7
Control 3 Temporal Ctx	17.9	Control (Path) 2 Parietal Ctx	33.2
Control (Path) 1 Temporal Ctx	67.8	Control (Path) 3 Parietal Ctx	8.7
Control (Path) 2 Temporal Ctx	52.5	Control (Path) 4 Parietal Ctx	59.5

Table GC. Panel 1.3D

Tissue Name	Rel. Exp.(%) Ag3088, Run 165552924	Tissue Name	Rel. Exp.(%) Ag3088, Run 165552924
Liver adenocarcinoma	37.1	Kidney (fetal)	11.7
Pancreas	18.9	Renal ca. 786-0	11.6
Pancreatic ca. CAPAN 2	31.2	Renal ca. A498	64.2
Adrenal gland	15.5	Renal ca. RXF 393	44.1
Thyroid	15.5	Renal ca. ACHN	25.7
Salivary gland	11.7	Renal ca. UO-31	36.6
Pituitary gland	10.3	Renal ca. TK-10	9.9
Brain (fetal)	46.7	Liver	12.1
Brain (whole)	70.2	Liver (fetal)	21.9
Brain (amygdala)	64.6	Liver ca. (hepatoblast) HepG2	52.1
Brain (cerebellum)	53.2	Lung	17.2
Brain (hippocampus)	77.4	Lung (fetal)	18.6
Brain (substantia nigra)	29.3	Lung ca. (small cell) LX-1	12.4
Brain (thalamus)	55.5	Lung ca. (small cell) NCI-H69	19.1
Cerebral Cortex	84.7	Lung ca. (s.cell var.) SHP-77	24.8
Spinal cord	19.2	Lung ca. (large cell) NCI-H460	18.0
glio/astro U87-MG	43.5	Lung ca. (non-sm. cell) A549	24.0
glio/astro U-118-MG	100.0	Lung ca. (non-s.cell) NCI-H23	6.0
astrocytoma SW1783	38.7	Lung ca. (non-s.cell) HOP-62	11.6
neuro*; met SK-N-AS	66.0	Lung ca. (non-s.cl) NCI-H522	6.2
astrocytoma SF-539	21.6	Lung ca. (squam.) SW 900	17.9
astrocytoma SNB-75	65.1	Lung ca. (squam.) NCI-H596	31.9
glioma SNB-19	40.3	Mammary gland	16.7
glioma U251	40.6	Breast ca.* (pl.ef) MCF-7	19.6
glioma SF-295	32.1	Breast ca.* (pl.ef) MDA-MB-231	80.1

Heart (fetal)	36.9	Breast ca.* (pl.ef) T47D	11.8
Heart	21.9	Breast ca. BT-549	44.8
Skeletal muscle (fetal)	14.4	Breast ca. MDA-N	12.6
Skeletal muscle	84.7	Ovary	22.2
Bone marrow	12.1	Ovarian ca. OVCAR-3	19.1
Thymus	6.7	Ovarian ca. OVCAR-4	85.3
Spleen	23.2	Ovarian ca. OVCAR-5	21.0
Lymph node	18.7	Ovarian ca. OVCAR-8	9.5
Colorectal	7.7	Ovarian ca. IGROV-1	5.7
Stomach	58.6	Ovarian ca.* (ascites) SK-OV-3	40.9
Small intestine	44.4	Uterus	19.9
Colon ca. SW480	18.9	Placenta	9.8
Colon ca.* SW620(SW480 met)	13.5	Prostate	16.6
Colon ca. HT29	12.1	Prostate ca.* (bone met)PC-3	87.7
Colon ca. HCT-116	19.1	Testis	23.8
Colon ca. CaCo-2	21.9	Melanoma Hs688(A).T	15.0
Colon ca. tissue(ODO3866)	16.3	Melanoma* (met) Hs688(B).T	12.4
Colon ca. HCC-2998	9.6	Melanoma UACC-62	31.6
Gastric ca.* (liver met) NCI-N87	41.2	Melanoma M14	36.6
Bladder	22.4	Melanoma LOX IMVI	24.1
Trachea	21.9	Melanoma* (met) SK-MEL-5	15.5
Kidney	24.0	Adipose	8.5

Table GD. Panel 2.2

Tissue Name	Rel. Exp.(%) Ag3088, Run 174268937	Tissue Name	Rel. Exp.(%) Ag3088, Run 174268937
Normal Colon	26.6	Kidney Margin (OD04348)	94.6
Colon cancer (OD06064)	14.9	Kidney malignant cancer (OD06204B)	12.2
Colon Margin (OD06064)	14.7	Kidney normal adjacent tissue (OD06204E)	29.1
Colon cancer (OD06159)	16.3	Kidney Cancer (OD04450-01)	42.9
Colon Margin (OD06159)	25.3	Kidney Margin (OD04450-03)	40.6
Colon cancer (OD06297-04)	11.5	Kidney Cancer 8120613	5.7
Colon Margin (OD06297-05)	15.8	Kidney Margin 8120614	52.1
CC Gr.2 ascend colon (ODO3921)	9.0	Kidney Cancer 9010320	15.5
CC Margin (ODO3921)	10.7	Kidney Margin 9010321	22.4
Colon cancer metastasis (OD06104)	5.8	Kidney Cancer 8120607	82.9
Lung Margin (OD06104)	17.4	Kidney Margin 8120608	35.4
Colon mets to lung (OD04451-01)	23.2	Normal Uterus	13.0

Lung Margin (OD04451-02)	19.1	Uterine Cancer 064011	12.3
Normal Prostate	22.4	Normal Thyroid	7.5
Prostate Cancer (OD04410)	7.8	Thyroid Cancer 064010	12.9
Prostate Margin (OD04410)	7.5	Thyroid Cancer A302152	27.9
Normal Ovary	48.3	Thyroid Margin A302153	7.1
Ovarian cancer (OD06283-03)	10.4	Normal Breast	15.0
Ovarian Margin (OD06283-07)	7.6	Breast Cancer (OD04566)	14.5
Ovarian Cancer 064008	11.9	Breast Cancer 1024	30.6
Ovarian cancer (OD06145)	11.7	Breast Cancer (OD04590-01)	60.3
Ovarian Margin (OD06145)	19.8	Breast Cancer Mets (OD04590-03)	25.2
Ovarian cancer (OD06455-03)	14.2	Breast Cancer Metastasis (OD04655-05)	55.1
Ovarian Margin (OD06455-07)	1.9	Breast Cancer 064006	20.3
Normal Lung	13.2	Breast Cancer 9100266	16.4
Invasive poor diff. lung adeno (ODO4945-01)	22.1	Breast Margin 9100265	7.2
Lung Margin (ODO4945-03)	13.5	Breast Cancer A209073	7.7
Lung Malignant Cancer (OD03126)	12.2	Breast Margin A2090734	17.8
Lung Margin (OD03126)	5.6	Breast cancer (OD06083)	29.5
Lung Cancer (OD05014A)	15.3	Breast cancer node metastasis (OD06083)	30.1
Lung Margin (OD05014B)	19.8	Normal Liver	50.3
Lung cancer (OD06081)	21.2	Liver Cancer 1026	27.7
Lung Margin (OD06081)	12.8	Liver Cancer 1025	80.1
Lung Cancer (OD04237-01)	5.5	Liver Cancer 6004-T	51.1
Lung Margin (OD04237-02)	23.5	Liver Tissue 6004-N	6.9
Ocular Melanoma Metastasis	16.4	Liver Cancer 6005-T	54.3
Ocular Melanoma Margin (Liver)	19.2	Liver Tissue 6005-N	100.0
Melanoma Metastasis	21.3	Liver Cancer 064003	62.4
Melanoma Margin (Lung)	6.9	Normal Bladder	19.8
Normal Kidney	12.8	Bladder Cancer 1023	10.0
Kidney Ca, Nuclear grade 2 (OD04338)	59.5	Bladder Cancer A302173	24.3
Kidney Margin (OD04338)	18.0	Normal Stomach	97.9
Kidney Ca Nuclear grade 1/2 (OD04339)	55.9	Gastric Cancer 9060397	13.5
Kidney Margin (OD04339)	26.4	Stomach Margin 9060396	41.2
Kidney Ca, Clear cell type (OD04340)	13.4	Gastric Cancer 9060395	25.9
Kidney Margin (OD04340)	28.9	Stomach Margin 9060394	37.4
Kidney Ca, Nuclear grade 3 (OD04348)	12.1	Gastric Cancer 064005	30.8

CNS_neurodegeneration_v1.0 Summary: Ag3088 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.3 for a discussion of the potential utility of this gene in treatment of central nervous system disorders.

Panel 1.3D Summary: Ag3088 This gene is widely expressed in many of the samples in this panel, with highest expression in a brain cancer U-118-MG cell line (CT = 26). This gene is also highly expressed in all the regions of the central nervous system, including the amygdala, cerebellum, hippocampus, substantia nigra, thalamus, cerebral cortex and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression. This gene codes for a homolog of epsin, which is involved in the phagocytosis of macromolecules, and interacts with Huntingtin-interacting protein. Therefore, this gene may play a critical role in the endocytosis of Huntingtin protein and the etiology of Huntington's disease. Downregulation of this gene or its protein product may be of therapeutic benefit in the treatment of Huntington's disease.

This gene is also expressed in many tissues with metabolic function, including pancreas, adrenal, thyroid, and pituitary glands, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

This gene is highly expressed in cell lines derived from melanoma, renal, breast, brain, ovarian, lung, colon, kidney, pancreatic and prostate cancers. Expression of this gene is higher in cancer cell lines when compared to corresponding normal tissues. Based on this expression profile, the expression of this gene may be used as a marker to detect these cancers. Furthermore, therapeutic modulation of this gene may be useful in the treatment of these cancers.

Panel 2.2 Summary: Ag3088 Highest expression of this gene is detected in normal liver tissue (CT = 27.3). In addition, the level of expression in some lung, breast, liver and kidney cancer tissue samples is higher than the corresponding adjacent control normal tissue.

The reverse appears to be true for colon, ovary and stomach tissue, where expression is slightly higher in normal tissue than the matched cancer tissues. Thus, based upon its profile, the expression of this gene may be used to distinguish between these cancers and the normal adjacent tissue. Please see panel 1.3 for further discussion on the utility of this gene.

5 H. NOV57a: GUANINE NUCLEOTIDE-BINDING PROTEIN GAMMA-7 SUBUNIT

Expression of gene NOV57a was assessed using the primer-probe set Ag4907, described in Table HA. Results of the RTQ-PCR runs are shown in Tables HB, HC and HD.

Table HA. Probe Name Ag4907

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gccgatctgctgaagttct-3'	19	127	366
Probe	TET-5'-aggccaagaatgaccccttctgt-3'-TAMRA	25	155	367
Reverse	5'-gcttcttctcctgaaggagtt-3'	22	199	368

10

Table HB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag4907, Run 214955039	Tissue Name	Rel. Exp.(%) Ag4907, Run 214955039
AD 1 Hippo	16.5	Control (Path) 3 Temporal Ctx	4.8
AD 2 Hippo	12.6	Control (Path) 4 Temporal Ctx	28.5
AD 3 Hippo	17.7	AD 1 Occipital Ctx	26.6
AD 4 Hippo	4.1	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	27.0
AD 6 Hippo	45.4	AD 4 Occipital Ctx	6.0
Control 2 Hippo	14.2	AD 5 Occipital Ctx	10.9
Control 4 Hippo	9.0	AD 6 Occipital Ctx	13.9
Control (Path) 3 Hippo	9.1	Control 1 Occipital Ctx	9.5
AD 1 Temporal Ctx	24.3	Control 2 Occipital Ctx	54.7
AD 2 Temporal Ctx	29.3	Control 3 Occipital Ctx	29.5
AD 3 Temporal Ctx	31.4	Control 4 Occipital Ctx	12.1
AD 4 Temporal Ctx	32.1	Control (Path) 1 Occipital Ctx	45.7
AD 5 Inf Temporal Ctx	81.2	Control (Path) 2 Occipital Ctx	23.5
AD 5 Sup Temporal Ctx	32.5	Control (Path) 3 Occipital Ctx	3.0
AD 6 Inf Temporal Ctx	17.9	Control (Path) 4 Occipital Ctx	22.5
AD 6 Sup Temporal Ctx	69.3	Control 1 Parietal Ctx	11.3
Control 1 Temporal Ctx	9.6	Control 2 Parietal Ctx	41.5
Control 2 Temporal Ctx	10.9	Control 3 Parietal Ctx	21.0
Control 3 Temporal Ctx	17.3	Control (Path) 1 Parietal Ctx	43.2

Control 3 Temporal Ctx	23.0	Control (Path) 2 Parietal Ctx	20.2
Control (Path) 1 Temporal Ctx	27.9	Control (Path) 3 Parietal Ctx	11.8
Control (Path) 2 Temporal Ctx	32.8	Control (Path) 4 Parietal Ctx	24.1

Table HC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4907, Run 228829503	Tissue Name	Rel. Exp.(%) Ag4907, Run 228829503
Adipose	5.5	Renal ca. TK-10	38.2
Melanoma* Hs688(A).T	7.6	Bladder	16.7
Melanoma* Hs688(B).T	4.1	Gastric ca. (liver met.) NCI-N87	100.0
Melanoma* M14	21.6	Gastric ca. KATO III	43.5
Melanoma* LOXIMVI	7.6	Colon ca. SW-948	11.7
Melanoma* SK-MEL-5	22.5	Colon ca. SW480	26.6
Squamous cell carcinoma SCC-4	5.9	Colon ca.* (SW480 met) SW620	28.7
Testis Pool	24.1	Colon ca. HT29	20.4
Prostate ca.* (bone met) PC-3	19.5	Colon ca. HCT-116	30.4
Prostate Pool	3.6	Colon ca. CaCo-2	33.9
Placenta	6.0	Colon cancer tissue	5.8
Uterus Pool	7.5	Colon ca. SW1116	4.7
Ovarian ca. OVCAR-3	43.8	Colon ca. Colo-205	11.3
Ovarian ca. SK-OV-3	42.3	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	15.6	Colon Pool	22.5
Ovarian ca. OVCAR-5	67.4	Small Intestine Pool	14.2
Ovarian ca. IGROV-1	13.0	Stomach Pool	13.6
Ovarian ca. OVCAR-8	15.3	Bone Marrow Pool	4.7
Ovary	14.4	Fetal Heart	6.0
Breast ca. MCF-7	39.0	Heart Pool	2.7
Breast ca. MDA-MB-231	37.6	Lymph Node Pool	13.9
Breast ca. BT 549	32.8	Fetal Skeletal Muscle	12.7
Breast ca. T47D	23.0	Skeletal Muscle Pool	5.7
Breast ca. MDA-N	20.9	Spleen Pool	1.6
Breast Pool	15.8	Thymus Pool	19.8
Trachea	4.9	CNS cancer (glio/astro) U87-MG	34.9
Lung	2.5	CNS cancer (glio/astro) U-118-MG	73.7
Fetal Lung	35.4	CNS cancer (neuro;met) SK-N-AS	50.3
Lung ca. NCI-N417	20.6	CNS cancer (astro) SF-539	11.9
Lung ca. LX-1	41.8	CNS cancer (astro) SNB-75	51.4
Lung ca. NCI-H146	26.2	CNS cancer (glio) SNB-19	13.9
Lung ca. SHP-77	63.7	CNS cancer (glio) SF-295	72.7
Lung ca. A549	17.1	Brain (Amygdala) Pool	5.5

Lung ca. NCI-H526	6.3	Brain (cerebellum)	47.3
Lung ca. NCI-H23	32.8	Brain (fetal)	31.0
Lung ca. NCI-H460	15.1	Brain (Hippocampus) Pool	12.9
Lung ca. HOP-62	15.0	Cerebral Cortex Pool	10.0
Lung ca. NCI-H522	93.3	Brain (Substantia nigra) Pool	10.7
Liver	0.9	Brain (Thalamus) Pool	7.2
Fetal Liver	9.2	Brain (whole)	12.8
Liver ca. HepG2	19.5	Spinal Cord Pool	2.3
Kidney Pool	31.0	Adrenal Gland	12.7
Fetal Kidney	42.3	Pituitary gland Pool	13.7
Renal ca. 786-0	21.9	Salivary Gland	6.6
Renal ca. A498	11.9	Thyroid (female)	7.8
Renal ca. ACHN	30.8	Pancreatic ca. CAPAN2	40.9
Renal ca. UO-31	9.4	Pancreas Pool	34.6 *

Table HD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4907, Run 223458613	Tissue Name	Rel. Exp.(%) Ag4907, Run 223458613
Secondary Th1 act	43.2	HUVEC IL-1beta	14.4
Secondary Th2 act	37.1	HUVEC IFN gamma	36.3
Secondary Tr1 act	7.4	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	6.6	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	8.4	HUVEC IL-11	35.6
Secondary Tr1 rest	5.8	Lung Microvascular EC none	32.8
Primary Th1 act	69.7	Lung Microvascular EC TNFalpha + IL-1beta	37.4
Primary Th2 act	96.6	Microvascular Dermal EC none	9.4
Primary Tr1 act	63.7	Microvascular Dermal EC TNFalpha + IL-1beta	16.6
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	24.5
Primary Th2 rest	7.6	Small airway epithelium none	26.8
Primary Tr1 rest	14.6	Small airway epithelium TNFalpha + IL-1beta	19.9
CD45RA CD4 lymphocyte act	46.3	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	58.2	Coronary artery SMC TNFalpha + IL-1beta	34.6
CD8 lymphocyte act	57.0	Astrocytes rest	48.3
Secondary CD8 lymphocyte rest	37.1	Astrocytes TNFalpha + IL-1beta	14.0
Secondary CD8 lymphocyte act	18.2	KU-812 (Basophil) rest	37.4
CD4 lymphocyte none	16.6	KU-812 (Basophil) PMA/ionomycin	27.5

2ry Th1/Th2/Tr1 _anti-CD95 CH11	13.5	CCD1106 (Keratinocytes) none	23.7
LAK cells rest	7.4	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	29.3
LAK cells IL-2	15.2	Liver cirrhosis	6.4
LAK cells IL-2+IL-12	9.2	NCI-H292 none	61.1
LAK cells IL-2+IFN gamma	17.3	NCI-H292 IL-4	48.6
LAK cells IL-2+ IL-18	34.4	NCI-H292 IL-9	100.0
LAK cells PMA/ionomycin	12.4	NCI-H292 IL-13	66.0
NK Cells IL-2 rest	22.4	NCI-H292 IFN gamma	54.3
Two Way MLR 3 day	6.0	HPAEC none	46.7
Two Way MLR 5 day	26.1	HPAEC TNF alpha + IL-1 beta	24.5
Two Way MLR 7 day	7.1	Lung fibroblast none	11.4
PBMC rest	25.7	Lung fibroblast TNF alpha + IL-1 beta	6.7
PBMC PWM	69.7	Lung fibroblast IL-4	20.4
PBMC PHA-L	20.7	Lung fibroblast IL-9	58.6
Ramos (B cell) none	34.6	Lung fibroblast IL-13	23.3
Ramos (B cell) ionomycin	27.4	Lung fibroblast IFN gamma	12.0
B lymphocytes PWM	44.4	Dermal fibroblast CCD1070 rest	30.6
B lymphocytes CD40L and IL-4	27.5	Dermal fibroblast CCD1070 TNF alpha	29.3
EOL-1 dbcAMP	41.2	Dermal fibroblast CCD1070 IL-1 beta	35.4
EOL-1 dbcAMP PMA/ionomycin	12.5	Dermal fibroblast IFN gamma	13.9
Dendritic cells none	12.3	Dermal fibroblast IL-4	49.3
Dendritic cells LPS	7.4	Dermal Fibroblasts rest	14.3
Dendritic cells anti-CD40	6.8	Neutrophils TNFa+LPS	0.0
Monocytes rest	27.4	Neutrophils rest	0.0
Monocytes LPS	7.9	Colon	0.0
Macrophages rest	7.9	Lung	17.4
Macrophages LPS	0.0	Thymus	7.5
HUVEC none	11.6	Kidney	40.1
HUVEC starved	22.1		

CNS_neurodegeneration_v1.0 Summary: Ag4907 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.5 for a discussion of the potential use of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.5 Summary: Ag4907 Highest expression of this gene is detected in gastric cancer NCI-N87 cell line (CT=31.3). Moderate levels of expression of

this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adrenal gland, thyroid, pituitary gland, skeletal muscle, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at low levels in most regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, and cerebral cortex. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Interestingly, this gene is expressed at much higher levels in fetal (CTs=32.8-34.8) when compared to adult lung and liver (CTs=36-38). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance growth or development of lung and liver in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung and liver related diseases.

Panel 4.1D Summary: Ag4907 Low levels of expression of this gene are detected mainly in IL-9 treated mucoepidermoid cell line NCI-H292. The expression of this gene in this mucoepidermoid cell line that is often used as a model for airway epithelium (NCI-H292 cells) suggests that this gene may be important in the proliferation or activation of airway epithelium. Therefore, therapeutics designed with the protein encoded by this gene may reduce or eliminate symptoms caused by inflammation in lung epithelia in chronic obstructive pulmonary disease, asthma, allergy, and emphysema.

I. NOV58a: Novel 2410017P07RIK Protein – Like Gene

Expression of gene NOV58a was assessed using the primer-probe set Ag4913, described in Table IA. Results of the RTQ-PCR runs are shown in Tables IB and IC.

Table IA. Probe Name Ag4913

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-attcaccagaatgaaccatct-3'	22	75	369
Probe	TET-5'-cagaattgccatcctgcaaactaga-3'-TAMRA	26	97	370
Reverse	5'-tggctatttgggctatgaagta-3'	22	140	371

5

Table IB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4913, Run 228829778	Tissue Name	Rel. Exp.(%) Ag4913, Run 228829778
Adipose	7.9	Renal ca. TK-10	13.0
Melanoma* Hs688(A).T	15.1	Bladder	13.8
Melanoma* Hs688(B).T	14.6	Gastric ca. (liver met.) NCI-N87	11.0
Melanoma* M14	8.0	Gastric ca. KATO III	27.2
Melanoma* LOXIMVI	19.5	Colon ca. SW-948	7.8
Melanoma* SK-MEL-5	13.4	Colon ca. SW480	20.4
Squamous cell carcinoma SCC-4	12.5	Colon ca.* (SW480 met) SW620	8.4
Testis Pool	21.0	Colon ca. HT29	10.7
Prostate ca.* (bone met) PC-3	11.3	Colon ca. HCT-116	55.5
Prostate Pool	7.6	Colon ca. CaCo-2	13.2
Placenta	2.0	Colon cancer tissue	13.8
Uterus Pool	8.4	Colon ca. SW1116	3.7
Ovarian ca. OVCAR-3	7.7	Colon ca. Colo-205	3.3
Ovarian ca. SK-OV-3	27.4	Colon ca. SW-48	5.2
Ovarian ca. OVCAR-4	6.6	Colon Pool	16.3
Ovarian ca. OVCAR-5	17.3	Small Intestine Pool	15.7
Ovarian ca. IGROV-1	6.7	Stomach Pool	10.4
Ovarian ca. OVCAR-8	6.7	Bone Marrow Pool	6.8
Ovary	7.1	Fetal Heart	9.3
Breast ca. MCF-7	20.3	Heart Pool	7.8
Breast ca. MDA-MB-231	39.8	Lymph Node Pool	21.8
Breast ca. BT 549	100.0	Fetal Skeletal Muscle	10.9
Breast ca. T47D	10.0	Skeletal Muscle Pool	24.5
Breast ca. MDA-N	9.7	Spleen Pool	10.5
Breast Pool	18.4	Thymus Pool	15.4
Trachea	8.9	CNS cancer (glio/astro) U87-MG	26.6
Lung	4.0	CNS cancer (glio/astro) U-118-MG	65.1

Fetal Lung	41.2	CNS cancer (neuro;met) SK-N-AS	14.2
Lung ca. NCI-N417	5.0	CNS cancer (astro) SF-539	19.3
Lung ca. LX-1	23.8	CNS cancer (astro) SNB-75	58.2
Lung ca. NCI-H146	11.9	CNS cancer (glio) SNB-19	8.7
Lung ca. SHP-77	32.1	CNS cancer (glio) SF-295	30.1
Lung ca. A549	16.3	Brain (Amygdala) Pool	7.2
Lung ca. NCI-H526	11.4	Brain (cerebellum)	39.2
Lung ca. NCI-H23	19.5	Brain (fetal)	18.7
Lung ca. NCI-H460	7.4	Brain (Hippocampus) Pool	7.3
Lung ca. HOP-62	5.6	Cerebral Cortex Pool	14.0
Lung ca. NCI-H522	36.6	Brain (Substantia nigra) Pool	6.4
Liver	0.8	Brain (Thalamus) Pool	15.8
Fetal Liver	31.6	Brain (whole)	6.5
Liver ca. HepG2	5.8	Spinal Cord Pool	7.8
Kidney Pool	25.3	Adrenal Gland	3.6
Fetal Kidney	30.8	Pituitary gland Pool	4.9
Renal ca. 786-0	26.8	Salivary Gland	1.1
Renal ca. A498	4.0	Thyroid (female)	4.5
Renal ca. ACHN	5.4	Pancreatic ca. CAPAN2	9.9
Renal ca. UO-31	12.9	Pancreas Pool	13.7

Table IC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4913, Run 223458616	Tissue Name	Rel. Exp.(%) Ag4913, Run 223458616
Secondary Th1 act	38.2	HUVEC IL-1beta	23.3
Secondary Th2 act	42.0	HUVEC IFN gamma	50.0
Secondary Tr1 act	38.2	HUVEC TNF alpha + IFN gamma	33.4
Secondary Th1 rest	10.0	HUVEC TNF alpha + IL4	20.9
Secondary Th2 rest	16.6	HUVEC IL-11	21.3
Secondary Tr1 rest	12.5	Lung Microvascular EC none	49.7
Primary Th1 act	12.3	Lung Microvascular EC TNFalpha + IL-1beta	35.8
Primary Th2 act	24.1	Microvascular Dermal EC none	32.3
Primary Tr1 act	23.3	Microvascular Dermal EC TNFalpha + IL-1beta	20.7
Primary Th1 rest	12.3	Bronchial epithelium TNFalpha + IL1beta	12.6
Primary Th2 rest	9.3	Small airway epithelium none	4.5
Primary Tr1 rest	19.6	Small airway epithelium TNFalpha + IL-1beta	8.5
CD45RA CD4 lymphocyte act	37.4	Coronary artery SMC rest	13.2

CD45RO CD4 lymphocyte act	42.3	Coronary artery SMC TNFalpha + IL-1beta	16.7
CD8 lymphocyte act	32.5	Astrocytes rest	7.1
Secondary CD8 lymphocyte rest	19.3	Astrocytes TNFalpha + IL-1beta	9.2
Secondary CD8 lymphocyte act	14.0	KU-812 (Basophil) rest	92.0
CD4 lymphocyte none	4.5	KU-812 (Basophil) PMA/ionomycin	100.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	24.7	CCD1106 (Keratinocytes) none	22.8
LAK cells rest	11.3	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	28.3
LAK cells IL-2	24.0	Liver cirrhosis	6.2
LAK cells IL-2+IL-12	5.6	NCI-H292 none	18.6
LAK cells IL-2+IFN gamma	15.1	NCI-H292 IL-4	25.2
LAK cells IL-2+ IL-18	14.9	NCI-H292 IL-9	38.4
LAK cells PMA/ionomycin	6.1	NCI-H292 IL-13	39.2
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	46.3
Two Way MLR 3 day	12.0	HPAEC none	18.9
Two Way MLR 5 day	17.4	HPAEC TNF alpha + IL-1 beta	24.7
Two Way MLR 7 day	13.9	Lung fibroblast none	29.9
PBMC rest	4.4	Lung fibroblast TNF alpha + IL-1 beta	16.3
PBMC PWM	12.7	Lung fibroblast IL-4	11.6
PBMC PHA-L	20.7	Lung fibroblast IL-9	12.9
Ramos (B cell) none	22.7	Lung fibroblast IL-13	14.2
Ramos (B cell) ionomycin	28.1	Lung fibroblast IFN gamma	37.1
B lymphocytes PWM	23.7	Dermal fibroblast CCD1070 rest	66.9
B lymphocytes CD40L and IL-4	21.8	Dermal fibroblast CCD1070 TNF alpha	76.3
EOL-1 dbcAMP	50.3	Dermal fibroblast CCD1070 IL-1 beta	37.4
EOL-1 dbcAMP PMA/ionomycin	23.7	Dermal fibroblast IFN gamma	2.0
Dendritic cells none	14.7	Dermal fibroblast IL-4	40.9
Dendritic cells LPS	9.9	Dermal Fibroblasts rest	24.1
Dendritic cells anti-CD40	10.9	Neutrophils TNFa+LPS	1.3
Monocytes rest	8.4	Neutrophils rest	6.4
Monocytes LPS	9.3	Colon	4.6
Macrophages rest	15.7	Lung	9.4
Macrophages LPS	5.8	Thymus	34.6
HUVEC none	22.7	Kidney	19.3
HUVEC starved	26.1		

General_screening_panel_v1.5 Summary: Ag4913 Highest expression of this gene is detected in breast cancer BT 549 cell line (CT=26.4). Moderate to high levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain

cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

5 Among tissues with metabolic or endocrine function, this gene is expressed at moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

10 Interestingly, this gene is expressed at much higher levels in fetal (CTs=27.7-28) when compared to adult lung and liver (CTs=31-33). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung and liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance growth or development of liver and lung in the fetus and thus may also act in a
15 regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung and liver related diseases.

 In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene
20 product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag4913 Highest expression of this gene is detected in basophils (Cts=29). This gene is expressed at high to moderate levels in a wide range of cell
25 types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for
30 these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional

therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

5 J. NOV59a: Novel FLJ20565 Like Gene

Expression of gene NOV59a was assessed using the primer-probe set Ag4914, described in Table JA. Results of the RTQ-PCR runs are shown in Table JB.

Table JA. Probe Name Ag4914

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ggggaagaaaagaaacaaagag-3'	22	646	372
Probe	TET-5'-ccccaacacagcctaaggccaag-3'-TAMRA	23	696	373
Reverse	5'-tcttaggctttccctcttagg-3'	22	722	374

10

Table JB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4914, Run 228839040	Tissue Name	Rel. Exp.(%) Ag4914, Run 228839040
Adipose	3.6	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.0	Bladder	0.0
Melanoma* Hs688(B).T	0.0	Gastric ca. (liver met.) NCI-N87	14.5
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	20.0	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	7.1
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	60.7
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	22.7
Ovarian ca. OVCAR-8	0.0	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0

Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	35.1	Thymus Pool	0.0
Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	24.1	CNS cancer (glio/astro) U-118-MG	74.7
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	14.4
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	69.7	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.0
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.0
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.0
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.0
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	0.0
Kidney Pool	18.8	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.0
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	100.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

General_screening_panel_v1.5 Summary: Ag4914 Low levels of expression of this gene are restricted to pancreatic cancer cell line (CT=34.5). Therefore, expression of this gene may be used to distinguish this sample from other samples in this panel and also as
5 diagnostic marker for detection of pancreatic cancer. Furthermore, therapeutic modulation of this gene may be useful in the treatment of this cancer.

K. NOV60a: CGI-27 Protein Like Gene

Expression of gene NOV60a was assessed using the primer-probe set Ag4915, described in Table KA. Results of the RTQ-PCR runs are shown in Tables KB and KC.

10 **Table KA. Probe Name Ag4915**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-agccactctgggaactggta-3'	20	31	375

Probe	TET-5'-ctcaggaccacagctgaatgcacag-3'-TAMRA	25	57	376
Reverse	5'-tacttgtagaagccaacctct-3'	22	84	377

Table KB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4915, Run 228839041	Tissue Name	Rel. Exp.(%) Ag4915, Run 228839041
Adipose	0.0	Renal ca. TK-10	30.1
Melanoma* Hs688(A).T	3.0	Bladder	11.8
Melanoma* Hs688(B).T	4.4	Gastric ca. (liver met.) NCI-N87	45.4
Melanoma* M14	15.7	Gastric ca. KATO III	9.3
Melanoma* LOXIMVI	26.6	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	10.2	Colon ca. SW480	7.5
Squamous cell carcinoma SCC-4	3.3	Colon ca.* (SW480 met) SW620	4.3
Testis Pool	100.0	Colon ca. HT29	5.8
Prostate ca.* (bone met) PC-3	6.0	Colon ca. HCT-116	86.5
Prostate Pool	0.0	Colon ca. CaCo-2	24.7
Placenta	0.0	Colon cancer tissue	9.2
Uterus Pool	0.0	Colon ca. SW1116	20.7
Ovarian ca. OVCAR-3	33.7	Colon ca. Colo-205	11.1
Ovarian ca. SK-OV-3	36.9	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	3.3	Colon Pool	0.0
Ovarian ca. OVCAR-5	71.2	Small Intestine Pool	8.4
Ovarian ca. IGROV-1	88.9	Stomach Pool	17.0
Ovarian ca. OVCAR-8	3.6	Bone Marrow Pool	15.8
Ovary	2.6	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	4.2	Lymph Node Pool	5.0
Breast ca. BT 549	59.9	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	61.1
Breast ca. MDA-N	0.0	Spleen Pool	0.0
Breast Pool	5.3	Thymus Pool	8.7
Trachea	0.0	CNS cancer (glio/astro) U87-MG	31.4
Lung	0.0	CNS cancer (glio/astro) U-118-MG	16.2
Fetal Lung	10.7	CNS cancer (neuro;met) SK-N-AS	7.6
Lung ca. NCI-N417	2.7	CNS cancer (astro) SF-539	3.7
Lung ca. LX-1	31.4	CNS cancer (astro) SNB-75	40.1
Lung ca. NCI-H146	92.0	CNS cancer (glio) SNB-19	25.2
Lung ca. SHP-77	3.0	CNS cancer (glio) SF-295	18.6
Lung ca. A549	8.0	Brain (Amygdala) Pool	7.9
Lung ca. NCI-H526	0.0	Brain (cerebellum)	18.7

Lung ca. NCI-H23	17.9	Brain (fetal)	42.9
Lung ca. NCI-H460	10.4	Brain (Hippocampus) Pool	7.2
Lung ca. HOP-62	12.4	Cerebral Cortex Pool	57.4
Lung ca. NCI-H522	8.8	Brain (Substantia nigra) Pool	24.8
Liver	0.0	Brain (Thalamus) Pool	49.7
Fetal Liver	0.0	Brain (whole)	48.6
Liver ca. HepG2	6.3	Spinal Cord Pool	18.6
Kidney Pool	7.3	Adrenal Gland	7.2
Fetal Kidney	22.4	Pituitary gland Pool	0.0
Renal ca. 786-0	20.9	Salivary Gland	14.1
Renal ca. A498	48.0	Thyroid (female)	0.0
Renal ca. ACHN	8.7	Pancreatic ca. CAPAN2	29.7
Renal ca. UO-31	22.5	Pancreas Pool	8.4

Table KC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4915, Run 223458640	Tissue Name	Rel. Exp.(%) Ag4915, Run 223458640
Secondary Th1 act	0.9	HUVEC IL-1beta	0.1
Secondary Th2 act	1.1	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.4	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	0.0	HUVEC IL-11	0.0
Secondary Tr1 rest	100.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	0.0
Primary Th2 act	0.2	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.1	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	0.2
Primary Th2 rest	0.0	Small airway epithelium none	0.5
Primary Tr1 rest	0.3	Small airway epithelium TNFalpha + IL-1beta	0.1
CD45RA CD4 lymphocyte act	0.3	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	0.6	Coronary artery SMC TNFalpha + IL-1beta	0.4
CD8 lymphocyte act	0.3	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.4
CD4 lymphocyte none	0.7	KU-812 (Basophil) PMA/ionomycin	0.4

2ry Th1/Th2/Tr1_anti-CD95 CH11	0.2	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	0.2	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.2
LAK cells IL-2	0.1	Liver cirrhosis	0.2
LAK cells IL-2+IL-12	0.6	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	0.1
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	0.3	NCI-H292 IL-13	0.3
NK Cells IL-2 rest	0.3	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	0.1	HPAEC none	0.1
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.3
Two Way MLR 7 day	0.5	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.3
PBMC PWM	0.5	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.2	Lung fibroblast IL-9	0.5
Ramos (B cell) none	0.1	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	0.1	Lung fibroblast IFN gamma	0.2
B lymphocytes PWM	0.5	Dermal fibroblast CCD1070 rest	0.1
B lymphocytes CD40L and IL-4	0.2	Dermal fibroblast CCD1070 TNF alpha	0.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	0.1	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.2
Dendritic cells anti-CD40	0.3	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.2	Colon	0.0
Macrophages rest	0.0	Lung	0.0
Macrophages LPS	0.0	Thymus	0.4
HUVEC none	0.0	Kidney	0.9
HUVEC starved	0.0		

General_screening_panel_v1.5 Summary: Ag4915 Highest expression of this gene is detected in testis (CT=34.3). Thus, expression of this gene could be used to differentiate between this sample and other samples on this panel and as a marker of testicular tissue.

- 5 Therapeutic modulation of the expression or function of this gene may be useful in the treatment of male infertility and hypogonadism.

In addition, low levels of expression of this gene is also seen in number of cancer cell lines derived from ovarian, lung and colon. Therefore, therapeutic modulation of this gene may be useful in the treatment of ovarian, lung and colon cancer. Furthermore, expression of

this gene may be used as diagnostic marker for the detection of colon, lung and ovarian cancers.

Panel 4.1D Summary: Ag4915 Moderate levels of expression of this gene is restricted to secondary Tr1 cells (CT=28.8). Thus, expression of this gene may be used to distinguish Tr1 cell from other samples used in this panel. Furthermore, expression of this gene in resting Tr1 cells suggest a role for this gene in T lymphocyte activation. Therefore, therapeutic modulation of this gene or its protein product may be useful in the treatment of T cell-mediated autoimmune and inflammatory diseases.

L. NOV64a: Ankyrin-repeat containing protein

Expression of gene NOV64a was assessed using the primer-probe set Ag4950, described in Table LA. Results of the RTQ-PCR runs are shown in Tables LB and LC.

Table LA. Probe Name Ag4950

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cttgagtgttgctgctaagca-3'	21	900	378
Probe	TET-5'-tccccgagaaagtgtagagcctta-3'-TAMRA	26	926	379
Reverse	5'-tccttttccatgggaaggt-3'	19	957	380

Table LB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4950, Run 228850857	Tissue Name	Rel. Exp.(%) Ag4950, Run 228850857
Adipose	4.1	Renal ca. TK-10	21.9
Melanoma* Hs688(A).T	3.6	Bladder	19.3
Melanoma* Hs688(B).T	4.0	Gastric ca. (liver met.) NCI-N87	50.7
Melanoma* M14	0.0	Gastric ca. KATO III	51.1
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	7.6
Melanoma* SK-MEL-5	15.4	Colon ca. SW480	14.9
Squamous cell carcinoma SCC-4	0.7	Colon ca. * (SW480 met) SW620	20.7
Testis Pool	100.0	Colon ca. HT29	19.1
Prostate ca. * (bone met) PC-3	4.8	Colon ca. HCT-116	3.9
Prostate Pool	2.9	Colon ca. CaCo-2	30.1
Placenta	8.8	Colon cancer tissue	21.0
Uterus Pool	2.1	Colon ca. SW1116	2.8
Ovarian ca. OVCAR-3	2.9	Colon ca. Colo-205	4.4
Ovarian ca. SK-OV-3	0.7	Colon ca. SW-48	0.7
Ovarian ca. OVCAR-4	0.6	Colon Pool	0.0

Ovarian ca. OVCAR-5	46.7	Small Intestine Pool	6.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	3.2
Ovarian ca. OVCAR-8	4.5	Bone Marrow Pool	3.2
Ovary	1.5	Fetal Heart	1.7
Breast ca. MCF-7	14.6	Heart Pool	1.2
Breast ca. MDA-MB-231	22.2	Lymph Node Pool	1.7
Breast ca. BT 549	0.5	Fetal Skeletal Muscle	0.0
Breast ca. T47D	11.4	Skeletal Muscle Pool	18.3
Breast ca. MDA-N	9.3	Spleen Pool	10.7
Breast Pool	1.3	Thymus Pool	31.9
Trachea	9.0	CNS cancer (glio/astro) U87-MG	3.5
Lung	0.0	CNS cancer (glio/astro) U-118-MG	41.5
Fetal Lung	12.9	CNS cancer (neuro;met) SK-N-AS	0.1
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	6.0
Lung ca. LX-1	7.3	CNS cancer (astro) SNB-75	15.1
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.2	CNS cancer (glio) SF-295	33.9
Lung ca. A549	20.4	Brain (Amygdala) Pool	5.8
Lung ca. NCI-H526	0.1	Brain (cerebellum)	0.1
Lung ca. NCI-H23	5.5	Brain (fetal)	2.0
Lung ca. NCI-H460	10.2	Brain (Hippocampus) Pool	2.1
Lung ca. HOP-62	3.6	Cerebral Cortex Pool	10.0
Lung ca. NCI-H522	69.3	Brain (Substantia nigra) Pool	6.0
Liver	0.0	Brain (Thalamus) Pool	8.7
Fetal Liver	0.0	Brain (whole)	0.0
Liver ca. HepG2	0.0	Spinal Cord Pool	4.9
Kidney Pool	3.4	Adrenal Gland	1.6
Fetal Kidney	4.4	Pituitary gland Pool	1.8
Renal ca. 786-0	10.4	Salivary Gland	1.6
Renal ca. A498	4.2	Thyroid (female)	1.8
Renal ca. ACHN	6.1	Pancreatic ca. CAPAN2	8.9
Renal ca. UO-31	10.6	Pancreas Pool	4.9

Table LC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4950, Run 223626816	Tissue Name	Rel. Exp.(%) Ag4950, Run 223626816
Secondary Th1 act	59.5	HUVEC IL-1 beta	1.0
Secondary Th2 act	79.6	HUVEC IFN gamma	0.0
Secondary Tr1 act	100.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	9.1	HUVEC TNF alpha + IL4	0.0

Secondary Th2 rest	3.0	HUVEC IL-11	1.8
Secondary Tr1 rest	15.0	Lung Microvascular EC none	3.6
Primary Th1 act	2.3	Lung Microvascular EC TNFalpha + IL-1beta	0.7
Primary Th2 act	11.5	Microvascular Dermal EC none	0.0
Primary Tr1 act	6.2	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	2.7	Bronchial epithelium TNFalpha + IL1beta	4.1
Primary Th2 rest	3.5	Small airway epithelium none	0.0
Primary Tr1 rest	12.9	Small airway epithelium TNFalpha + IL-1beta	0.0
CD45RA CD4 lymphocyte act	4.3	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	16.4	Coronary artery SMC TNFalpha + IL-1beta	0.0
CD8 lymphocyte act	18.9	Astrocytes rest	0.0
Secondary CD8 lymphocyte rest	22.5	Astrocytes TNFalpha + IL-1beta	0.0
Secondary CD8 lymphocyte act	2.8	KU-812 (Basophil) rest	4.0
CD4 lymphocyte none	20.2	KU-812 (Basophil) PMA/ionomycin	2.0
2ry Th1/Th2/Tr1 _anti-CD95 CH11	11.0	CCD1106 (Keratinocytes) none	0.0
LAK cells rest	15.4	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	0.0
LAK cells IL-2	28.3	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	28.7	NCI-H292 none	31.2
LAK cells IL-2+IFN gamma	26.4	NCI-H292 IL-4	31.6
LAK cells IL-2+ IL-18	24.1	NCI-H292 IL-9	41.5
LAK cells PMA/ionomycin	9.7	NCI-H292 IL-13	15.0
NK Cells IL-2 rest	32.5	NCI-H292 IFN gamma	35.8
Two Way MLR 3 day	20.9	HPAEC none	0.7
Two Way MLR 5 day	6.5	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	20.2	Lung fibroblast none	0.0
PBMC rest	10.8	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	30.8	Lung fibroblast IL-4	0.0
PBMC PHA-L	3.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	1.1	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	2.3	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	5.3	Dermal fibroblast CCD1070 rest	0.0
B lymphocytes CD40L and IL-4	3.0	Dermal fibroblast CCD1070 TNF alpha	1.4
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	1.4
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	6.1	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	1.5
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0

Monocytes rest	0.0	Neutrophils rest	1.3
Monocytes LPS	0.0	Colon	0.0
Macrophages rest	1.5	Lung	6.3
Macrophages LPS	0.0	Thymus	52.5
HUVEC none	0.0	Kidney	9.2
HUVEC starved	0.0		

General_screening_panel_v1.5 Summary: Ag4950 Highest expression of this gene is detected in testis (CT=29.3). Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of male infertility and hypogonadism.

5 Moderate to low levels of expression of this gene is also seen in number of cancer cell lines derived from pancreatic, gastric, colon, lung, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric,
10 colon, lung, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

 Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, skeletal muscle, and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the
15 treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

 In addition, this gene is expressed at low levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease,
20 Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

 Interestingly, this gene is expressed at much higher levels in fetal (CT=32) when compared to adult lung (CT=40). This observation suggests that expression of this gene can be used to distinguish fetal from adult lung. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance lung growth or
25 development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of lung related diseases.

Panel 4.1D Summary: Ag4950 Highest expression of this gene is detected in activated secondary Tr1 cells (CT=32). In addition, moderate to low levels of expression of this gene is also seen in activated polarized T cells, memory T cells, LAK cells, activated PBMC, mucoepidermoid NCI-H292 cells, and thymus. Expression of this gene is upregulated in activated secondary polarized T cells as well as in PBMC cells. Thus, this gene may be involved in activation T and B cells. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

M. NOV65a: MULTIDOMAIN PRESYNAPTIC CYTOMATRIX PROTEIN PICCOLO

Expression of gene NOV65a was assessed using the primer-probe set Ag4951, described in Table MA. Results of the RTQ-PCR runs are shown in Tables MB and MC.

Table MA. Probe Name Ag4951

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgcactgaatgcaagaatca-3'	20	3133	381
Probe	TET-5'-tctctgtggatttaaccctacaccaca-3'-TAMRA	27	3162	382
Reverse	5'-agccattcttgaatctcagtc-3'	22	3191	383

Table MB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag4951, Run 249286336	Tissue Name	Rel. Exp.(%) Ag4951, Run 249286336
AD 1 Hippo	8.4	Control (Path) 3 Temporal Ctx	3.9
AD 2 Hippo	25.7	Control (Path) 4 Temporal Ctx	57.4
AD 3 Hippo	8.1	AD 1 Occipital Ctx	18.4
AD 4 Hippo	6.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	77.9	AD 3 Occipital Ctx	6.0
AD 6 Hippo	53.2	AD 4 Occipital Ctx	25.5
Control 2 Hippo	43.2	AD 5 Occipital Ctx	53.6
Control 4 Hippo	3.4	AD 6 Occipital Ctx	24.1
Control (Path) 3 Hippo	3.8	Control 1 Occipital Ctx	1.3
AD 1 Temporal Ctx	17.4	Control 2 Occipital Ctx	42.9

AD 2 Temporal Ctx	31.4	Control 3 Occipital Ctx	27.0
AD 3 Temporal Ctx	6.6	Control 4 Occipital Ctx	3.4
AD 4 Temporal Ctx	22.1	Control (Path) 1 Occipital Ctx	98.6
AD 5 Inf Temporal Ctx	62.4	Control (Path) 2 Occipital Ctx	17.8
AD 5 Sup Temporal Ctx	34.4	Control (Path) 3 Occipital Ctx	0.8
AD 6 Inf Temporal Ctx	52.1	Control (Path) 4 Occipital Ctx	31.9
AD 6 Sup Temporal Ctx	58.6	Control 1 Parietal Ctx	4.7
Control 1 Temporal Ctx	3.5	Control 2 Parietal Ctx	36.3
Control 2 Temporal Ctx	32.8	Control 3 Parietal Ctx	13.3
Control 3 Temporal Ctx	27.4	Control (Path) 1 Parietal Ctx	92.0
Control 3 Temporal Ctx	4.0	Control (Path) 2 Parietal Ctx	28.5
Control (Path) 1 Temporal Ctx	100.0	Control (Path) 3 Parietal Ctx	2.5
Control (Path) 2 Temporal Ctx	54.3	Control (Path) 4 Parietal Ctx	65.5

Table MC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4951, Run 223626818	Tissue Name	Rel. Exp.(%) Ag4951, Run 223626818
Secondary Th1 act	0.0	HUVEC IL-1beta	3.5
Secondary Th2 act	0.0	HUVEC IFN gamma	0.4
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.9
Secondary Th2 rest	0.0	HUVEC IL-11	1.6
Secondary Tr1 rest	0.0	Lung Microvascular EC none	9.1
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	2.3
Primary Th2 act	1.2	Microvascular Dermal EC none	11.6
Primary Tr1 act	0.0	Microvascular Dermal EC TNFalpha + IL-1beta	3.2
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	40.3
Primary Th2 rest	0.0	Small airway epithelium none	3.3
Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	17.3
CD45RA CD4 lymphocyte act	0.0	Coronary artery SMC rest	0.7
CD45RO CD4 lymphocyte act	0.0	Coronary artery SMC TNFalpha + IL-1beta	0.8
CD8 lymphocyte act	0.0	Astrocytes rest	1.7
Secondary CD8 lymphocyte rest	0.0	Astrocytes TNFalpha + IL-1beta	2.1
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	0.0	KU-812 (Basophil) PMA/ionomycin	0.0

2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	44.4
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	36.6
LAK cells IL-2	0.0	Liver cirrhosis	25.0
LAK cells IL-2+IL-12	0.0	NCI-H292 none	14.4
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	18.7
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	39.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	26.8
NK Cells IL-2 rest	0.0	NCI-H292 IFN gamma	23.3
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	0.4	Lung fibroblast none	0.0
PBMC rest	0.0	Lung fibroblast TNF alpha + IL-1 beta	0.7
PBMC PWM	0.0	Lung fibroblast IL-4	0.0
PBMC PHA-L	0.0	Lung fibroblast IL-9	0.0
Ramos (B cell) none	42.3	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	47.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	4.6
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	1.0
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	0.0
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.8
Dendritic cells none	0.0	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	5.7
Macrophages rest	0.8	Lung	17.8
Macrophages LPS	0.0	Thymus	38.2
HUVEC none	2.2	Kidney	100.0
HUVEC starved	2.5		

CNS_neurodegeneration_v1.0 Summary: Ag4951 Expression of this gene is ubiquitous throughout the samples in this panel, with highest expression in the temporal cortex of a control patient with pathological condition (CT=26.4). While no association
5 between the expression of this gene and the presence of Alzheimer's disease is detected in this panel, these results confirm the expression of this gene in areas that degenerate in Alzheimer's disease, including the cortex, hippocampus, amygdala and thalamus. Expression of this gene in brain suggests that this gene may play a role in central nervous system

disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag4951 Highest expression of this gene is detected in kidney (CT=30.1). In addition, moderate to low levels of expression of this gene is also seen in colon, lung, thymus, Ramos B cells, lung microvascular endothelial cells, cytokine activated small airway epithelium, keratinocytes, mucoepidermoid NCI-H292 cells and liver cirrhosis samples. Therefore, therapeutic modulation of this gene may be useful in the treatment of autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, osteoarthritis and liver cirrhosis.

10 **N. NOV66a and NOV66b: human ortholog of rat CYTOSOLIC SORTING PROTEIN PACS-1A**

Expression of gene NOV66a and variant NOV66b was assessed using the primer-probe sets Ag4956 and Ag4960, described in Tables NA and NB. Results of the RTQ-PCR runs are shown in Tables NC and ND.

15 **Table NA. Probe Name Ag4956**

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaagatctccggaaagtgaaga-3'	22	923	384
Probe	TET-5'-cccggaggaaactaacctcaacctct-3'-TAMRA	26	948	385
Reverse	5'-gatgttaggttccttgtgatg-3'	22	976	386

Table NB. Probe Name Ag4960

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gaagatctccggaaagtgaaga-3'	22	923	387
Probe	TET-5'-cccggaggaaactaacctcaacctct-3'-TAMRA	26	948	388
Reverse	5'-gatgttaggttccttgtgatg-3'	22	976	389

20

Table NC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4956, Run 228886975	Rel. Exp.(%) Ag4960, Run 228886991	Tissue Name	Rel. Exp.(%) Ag4956, Run 228886975	Rel. Exp.(%) Ag4960, Run 228886991
Adipose	15.1	2.2	Renal ca. TK-10	26.6	27.7
Melanoma* Hs688(A).T	79.6	84.1	Bladder	30.4	2.5

Melanoma* Hs688(B).T	84.7	81.8	Gastric ca. (liver met.) NCI-N87	56.6	45.1
Melanoma* M14	32.3	31.2	Gastric ca. KATO III	44.4	41.5
Melanoma* LOXIMVI	45.4	40.6	Colon ca. SW-948	4.7	4.5
Melanoma* SK- MEL-5	40.6	46.0	Colon ca. SW480	48.0	43.5
Squamous cell carcinoma SCC-4	33.7	32.5	Colon ca.* (SW480 met) SW620	16.2	17.7
Testis Pool	23.2	26.2	Colon ca. HT29	8.4	8.0
Prostate ca.* (bone met) PC-3	32.8	34.2	Colon ca. HCT-116	41.8	38.4
Prostate Pool	9.9	10.4	Colon ca. CaCo-2	35.1	30.6
Placenta	20.2	17.2	Colon cancer tissue	16.5	15.3
Uterus Pool	12.0	13.2	Colon ca. SW1116	8.4	7.4
Ovarian ca. OVCAR- 3	33.7	33.0	Colon ca. Colo-205	1.8	1.5
Ovarian ca. SK-OV-3	67.8	63.3	Colon ca. SW-48	2.3	1.8
Ovarian ca. OVCAR- 4	32.1	29.9	Colon Pool	22.4	22.4
Ovarian ca. OVCAR- 5	37.4	38.7	Small Intestine Pool	12.9	10.7
Ovarian ca. IGROV-I	18.7	14.5	Stomach Pool	9.7	9.9
Ovarian ca. OVCAR- 8	17.7	15.9	Bone Marrow Pool	9.0	6.7
Ovary	22.8	20.6	Fetal Heart	19.2	19.5
Breast ca. MCF-7	42.3	42.6	Heart Pool	11.2	9.4
Breast ca. MDA-MB- 231	97.3	100.0	Lymph Node Pool	24.8	20.9
Breast ca. BT 549	59.9	58.6	Fetal Skeletal Muscle	6.8	7.6
Breast ca. T47D	12.2	15.9	Skeletal Muscle Pool	13.8	13.8
Breast ca. MDA-N	12.4	14.9	Spleen Pool	11.8	10.7
Breast Pool	23.3	21.2	Thymus Pool	17.3	17.2
Trachea	33.0	34.2	CNS cancer (glio/astro) U87-MG	62.9	56.3
Lung	4.9	4.7	CNS cancer (glio/astro) U-118-MG	74.7	74.7
Fetal Lung	30.8	31.9	CNS cancer (neuro;met) SK-N-AS	69.7	66.4
Lung ca. NCI-N417	5.6	5.0	CNS cancer (astro) SF- 539	20.0	20.6
Lung ca. LX-I	20.4	20.0	CNS cancer (astro) SNB- 75	79.0	71.2
Lung ca. NCI-H146	6.9	7.5	CNS cancer (glio) SNB- 19	15.0	13.3
Lung ca. SHP-77	17.9	15.6	CNS cancer (glio) SF-295	43.5	43.5

Lung ca. A549	81.2	78.5	Brain (Amygdala) Pool	14.2	8.0
Lung ca. NCI-H526	7.4	9.0	Brain (cerebellum)	45.4	29.9
Lung ca. NCI-H23	19.5	24.5	Brain (fetal)	100.0	77.4
Lung ca. NCI-H460	11.5	12.6	Brain (Hippocampus) Pool	15.6	17.0
Lung ca. HOP-62	26.2	22.5	Cerebral Cortex Pool	23.0	18.4
Lung ca. NCI-H522	38.4	39.2	Brain (Substantia nigra) Pool	16.6	12.7
Liver	1.9	1.5	Brain (Thalamus) Pool	25.5	20.2
Fetal Liver	12.7	12.2	Brain (whole)	48.3	31.2
Liver ca. HepG2	13.2	12.9	Spinal Cord Pool	11.0	9.5
Kidney Pool	25.5	25.5	Adrenal Gland	17.3	12.9
Fetal Kidney	11.0	8.7	Pituitary gland Pool	4.1	3.9
Renal ca. 786-0	35.6	34.9	Salivary Gland	57.0	42.0
Renal ca. A498	10.2	8.7	Thyroid (female)	10.3	11.2
Renal ca. ACHN	37.1	39.2	Pancreatic ca. CAPAN2	35.6	36.3
Renal ca. UO-31	86.5	79.0	Pancreas Pool	21.0	21.6

Table ND. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4956, Run 223629803	Rel. Exp.(%) Ag4960, Run 223629875	Tissue Name	Rel. Exp.(%) Ag4956, Run 223629803	Rel. Exp.(%) Ag4960, Run 223629875
Secondary Th1 act	39.0	39.2	HUVEC IL-1beta	32.1	32.5
Secondary Th2 act	42.0	31.4	HUVEC IFN gamma	38.2	40.1
Secondary Tr1 act	28.9	10.0	HUVEC TNF alpha + IFN gamma	27.9	25.2
Secondary Th1 rest	36.1	24.3	HUVEC TNF alpha + IL4	26.8	25.9
Secondary Th2 rest	55.1	26.6	HUVEC IL-11	24.7	24.8
Secondary Tr1 rest	61.6	16.4	Lung Microvascular EC none	67.8	69.7
Primary Th1 act	27.5	14.3	Lung Microvascular EC TNFalpha + IL-1beta	51.8	53.2
Primary Th2 act	82.4	45.4	Microvascular Dermal EC none	43.5	45.1
Primary Tr1 act	56.3	24.5	Microvascular Dermal EC TNFalpha + IL- 1beta	38.7	40.9
Primary Th1 rest	48.0	31.4	Bronchial epithelium TNFalpha + IL1beta	32.8	35.1
Primary Th2 rest	74.2	49.3	Small airway epithelium none	15.6	15.6

Primary Tr1 rest	73.7	41.2	Small airway epithelium TNFalpha + IL-1 beta	24.5	26.6
CD45RA CD4 lymphocyte act	49.3	50.0	Coronary artery SMC rest	63.3	65.1
CD45RO CD4 lymphocyte act	77.9	83.5	Coronary artery SMC TNFalpha + IL-1 beta	55.9	66.0
CD8 lymphocyte act	44.4	51.1	Astrocytes rest	28.9	39.8
Secondary CD8 lymphocyte rest	52.1	54.0	Astrocytes TNFalpha + IL-1 beta	31.4	39.5
Secondary CD8 lymphocyte act	30.1	32.1	KU-812 (Basophil) rest	16.3	19.6
CD4 lymphocyte none	38.4	47.6	KU-812 (Basophil) PMA/ionomycin	34.2	38.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	90.8	92.7	CCD1106 (Keratinocytes) none	33.9	36.9
LAK cells rest	37.9	35.1	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	42.0	43.2
LAK cells IL-2	52.9	51.4	Liver cirrhosis	14.9	16.2
LAK cells IL-2+IL-12	27.0	25.2	NCI-H292 none	30.1	29.1
LAK cells IL-2+IFN gamma	40.6	43.2	NCI-H292 IL-4	47.6	50.0
LAK cells IL-2+ IL-18	48.3	45.4	NCI-H292 IL-9	61.6	58.6
LAK cells PMA/ionomycin	16.8	15.1	NCI-H292 IL-13	59.0	53.2
NK Cells IL-2 rest	65.5	59.5	NCI-H292 IFN gamma	53.6	67.4
Two Way MLR 3 day	35.6	38.2	HPAEC none	26.4	24.5
Two Way MLR 5 day	28.9	32.3	HPAEC TNF alpha + IL-1 beta	33.2	33.7
Two Way MLR 7 day	36.3	45.4	Lung fibroblast none	45.1	47.3
PBMC rest	34.2	42.0	Lung fibroblast TNF alpha + IL-1 beta	26.6	30.6
PBMC PWM	31.0	26.2	Lung fibroblast IL-4	46.7	37.1
PBMC PHA-L	56.6	50.0	Lung fibroblast IL-9	47.3	50.3
Ramos (B cell) none	8.1	8.1	Lung fibroblast IL-13	33.7	41.5
Ramos (B cell) ionomycin	8.8	8.2	Lung fibroblast IFN gamma	53.6	57.8
B lymphocytes PWM	34.6	38.7	Dermal fibroblast CCD1070 rest	54.3	50.0
B lymphocytes CD40L and IL-4	63.7	61.6	Dermal fibroblast CCD1070 TNF alpha	100.0	100.0
EOL-1 dbcAMP	14.6	14.1	Dermal fibroblast CCD1070 IL-1 beta	35.6	45.4
EOL-1 dbcAMP PMA/ionomycin	18.6	20.7	Dermal fibroblast IFN gamma	38.4	32.1
Dendritic cells none	53.6	46.3	Dermal fibroblast IL-4	51.1	52.9

Dendritic cells LPS	31.2	28.9	Dermal Fibroblasts rest	37.6	41.5
Dendritic cells anti-CD40	43.5	57.8	Neutrophils TNFa+LPS	25.7	27.7
Monocytes rest	39.2	34.2	Neutrophils rest	64.6	61.1
Monocytes LPS	9.2	13.0	Colon	9.2	6.8
Macrophages rest	27.2	42.0	Lung	27.9	24.8
Macrophages LPS	12.9	10.0	Thymus	32.5	31.9
HUVEC none	18.6	19.9	Kidney	16.6	16.4
HUVEC starved	27.9	30.8			

General_screening_panel_v1.5 Summary: Ag4956/Ag4960 Two experiments with same probe and primer sets are in excellent agreement with highest expression of this gene seen in fetal brain and breast cancer MDA-MB-231 cell line (CTs=25). Moderate to high levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at high to moderate levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

In addition, this gene is expressed at high levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag4956/Ag4960 Two experiments with same probe and primer sets are in excellent agreement with highest expression of this gene seen in TNF alpha treated dermal fibroblast (CTs=27-27.5). This gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte,

and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

O. NOV67a: Formin 2

Expression of gene NOV67a was assessed using the primer-probe set Ag4959, described in Table OA. Results of the RTQ-PCR runs are shown in Tables OB and OC.

Table OA. Probe Name Ag4959

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgcattgtgatggttttgtatg-3'	22	5198	390
Probe	TET-5'-tccattacttttaatgatcttcgttgca-3'-TAMRA	28	5235	391
Reverse	5'-aaattcctcccctttatccacaa-3'	22	5271	392

Table OB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag4959, Run 224735164	Tissue Name	Rel. Exp.(%) Ag4959, Run 224735164
AD 1 Hippo	2.9	Control (Path) 3 Temporal Ctx	4.7
AD 2 Hippo	3.9	Control (Path) 4 Temporal Ctx	19.9
AD 3 Hippo	1.1	AD 1 Occipital Ctx	9.7
AD 4 Hippo	2.7	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	46.0	AD 3 Occipital Ctx	1.2
AD 6 Hippo	100.0	AD 4 Occipital Ctx	4.4
Control 2 Hippo	6.4	AD 5 Occipital Ctx	9.9
Control 4 Hippo	2.5	AD 6 Occipital Ctx	15.7
Control (Path) 3 Hippo	1.2	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	5.7	Control 2 Occipital Ctx	52.1
AD 2 Temporal Ctx	7.4	Control 3 Occipital Ctx	5.5
AD 3 Temporal Ctx	0.0	Control 4 Occipital Ctx	1.6

AD 4 Temporal Ctx	3.5	Control (Path) 1 Occipital Ctx	41.5
AD 5 Inf Temporal Ctx	48.0	Control (Path) 2 Occipital Ctx	2.9
AD 5 Sup Temporal Ctx	17.6	Control (Path) 3 Occipital Ctx	0.0
AD 6 Inf Temporal Ctx	27.7	Control (Path) 4 Occipital Ctx	5.8
AD 6 Sup Temporal Ctx	26.6	Control 1 Parietal Ctx	1.4
Control 1 Temporal Ctx	1.0	Control 2 Parietal Ctx	17.6
Control 2 Temporal Ctx	19.2	Control 3 Parietal Ctx	4.6
Control 3 Temporal Ctx	5.7	Control (Path) 1 Parietal Ctx	46.0
Control 4 Temporal Ctx	2.0	Control (Path) 2 Parietal Ctx	4.8
Control (Path) 1 Temporal Ctx	22.5	Control (Path) 3 Parietal Ctx	2.2
Control (Path) 2 Temporal Ctx	25.9	Control (Path) 4 Parietal Ctx	23.5

Table OC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4959, Run 228886990	Tissue Name	Rel. Exp.(%) Ag4959, Run 228886990
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	2.3	Bladder	0.0
Melanoma* Hs688(B).T	5.5	Gastric ca. (liver met.) NCI-N87	0.0
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	5.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	100.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca. * (SW480 met) SW620	0.0
Testis Pool	0.0	Colon ca. HT29	0.0
Prostate ca. * (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	0.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.0
Ovarian ca. OVCAR-8	0.4	Bone Marrow Pool	0.0
Ovary	0.0	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.0
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	0.0
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.6	Spleen Pool	0.0
Breast Pool	0.0	Thymus Pool	0.0

Trachea	0.0	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	1.3
Fetal Lung	0.0	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	0.0	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	0.4	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.8
Lung ca. NCI-H526	0.0	Brain (cerebellum)	4.7
Lung ca. NCI-H23	0.0	Brain (fetal)	1.2
Lung ca. NCI-H460	0.7	Brain (Hippocampus) Pool	1.2
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	2.3
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	1.0
Liver	0.0	Brain (Thalamus) Pool	2.3
Fetal Liver	0.0	Brain (whole)	2.3
Liver ca. HepG2	0.0	Spinal Cord Pool	0.8
Kidney Pool	0.0	Adrenal Gland	0.0
Fetal Kidney	0.0	Pituitary gland Pool	0.4
Renal ca. 786-0	0.0	Salivary Gland	0.0
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.0
Renal ca. UO-31	0.0	Pancreas Pool	0.0

CNS_neurodegeneration_v1.0 Summary: Ag4959 Low levels of expression of this gene is seen throughout the samples in this panel, with highest expression in the hippocampus of a patient with Alzheimer's disease (CT=33.3). While no association between the expression of this gene and the presence of Alzheimer's disease is detected in this panel, these results confirm the expression of this gene in areas that degenerate in Alzheimer's disease, including the cortex, hippocampus, amygdala and thalamus.

General_screening_panel_v1.5 Summary: Ag4959 Moderate levels of expression of this gene are restricted to melanoma SK-MEL-5 cell line (CT=31.7). Therefore, expression of this gene may be used to distinguish this sample from other samples used in this panel and also as marker to detect the presence of melanoma. Furthermore, therapeutic modulation of this gene may be useful in the treatment of melanoma.

P. NOV69a: F-box domain containing protein

Expression of gene NOV69a was assessed using the primer-probe set Ag4961, described in Table PA. Results of the RTQ-PCR runs are shown in Tables PB, PC and PD.

Table PA. Probe Name Ag4961

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cagaggtgcccttagcttact-3'	22	1671	393
Probe	TET-5'-ttctcagcgtagattttgtccatcaa-3'-TAMRA	26	1706	394
Reverse	5'-caaatggcggcatgtataatc-3'	22	1745	395

Table PB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag4961, Run 224735209	Tissue Name	Rel. Exp.(%) Ag4961, Run 224735209
AD 1 Hippo	14.6	Control (Path) 3 Temporal Ctx	7.4
AD 2 Hippo	33.9	Control (Path) 4 Temporal Ctx	32.1
AD 3 Hippo	10.6	AD 1 Occipital Ctx	18.0
AD 4 Hippo	4.9	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	100.0	AD 3 Occipital Ctx	12.7
AD 6 Hippo	77.9	AD 4 Occipital Ctx	23.2
Control 2 Hippo	24.1	AD 5 Occipital Ctx	32.8
Control 4 Hippo	17.9	AD 6 Occipital Ctx	34.9
Control (Path) 3 Hippo	11.5	Control 1 Occipital Ctx	8.4
AD 1 Temporal Ctx	27.0	Control 2 Occipital Ctx	38.4
AD 2 Temporal Ctx	39.0	Control 3 Occipital Ctx	19.8
AD 3 Temporal Ctx	8.1	Control 4 Occipital Ctx	9.5
AD 4 Temporal Ctx	25.5	Control (Path) 1 Occipital Ctx	84.7
AD 5 Inf Temporal Ctx	88.9	Control (Path) 2 Occipital Ctx	12.9
AD 5 Sup Temporal Ctx	43.2	Control (Path) 3 Occipital Ctx	6.2
AD 6 Inf Temporal Ctx	57.0	Control (Path) 4 Occipital Ctx	17.0
AD 6 Sup Temporal Ctx	80.7	Control 1 Parietal Ctx	9.6
Control 1 Temporal Ctx	11.7	Control 2 Parietal Ctx	52.5
Control 2 Temporal Ctx	36.1	Control 3 Parietal Ctx	15.1
Control 3 Temporal Ctx	13.6	Control (Path) 1 Parietal Ctx	66.9
Control 3 Temporal Ctx	13.0	Control (Path) 2 Parietal Ctx	25.7
Control (Path) 1 Temporal Ctx	54.7	Control (Path) 3 Parietal Ctx	6.2
Control (Path) 2 Temporal Ctx	27.5	Control (Path) 4 Parietal Ctx	44.1

5

Table PC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4961, Run 228903662	Tissue Name	Rel. Exp.(%) Ag4961, Run 228903662
-------------	---	-------------	---

Adipose	7.4	Renal ca. TK-10	20.0
Melanoma* Hs688(A).T	40.6	Bladder	12.8
Melanoma* Hs688(B).T	40.9	Gastric ca. (liver met.) NCI-N87	12.6
Melanoma* M14	10.1	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	20.7	Colon ca. SW-948	5.6
Melanoma* SK-MEL-5	18.7	Colon ca. SW480	23.3
Squamous cell carcinoma SCC-4	12.0	Colon ca.* (SW480 met) SW620	29.7
Testis Pool	11.5	Colon ca. HT29	5.3
Prostate ca.* (bone met) PC-3	25.0	Colon ca. HCT-116	50.0
Prostate Pool	12.3	Colon ca. CaCo-2	29.7
Placenta	1.9	Colon cancer tissue	8.8
Uterus Pool	15.1	Colon ca. SW1116	2.7
Ovarian ca. OVCAR-3	13.7	Colon ca. Colo-205	0.9
Ovarian ca. SK-OV-3	54.3	Colon ca. SW-48	5.6
Ovarian ca. OVCAR-4	4.3	Colon Pool	28.7
Ovarian ca. OVCAR-5	15.0	Small Intestine Pool	21.2
Ovarian ca. IGROV-1	20.4	Stomach Pool	9.9
Ovarian ca. OVCAR-8	12.9	Bone Marrow Pool	6.1
Ovary	11.9	Fetal Heart	7.2
Breast ca. MCF-7	16.3	Heart Pool	11.8
Breast ca. MDA-MB-231	24.1	Lymph Node Pool	27.4
Breast ca. BT 549	60.3	Fetal Skeletal Muscle	4.5
Breast ca. T47D	2.9	Skeletal Muscle Pool	33.2
Breast ca. MDA-N	9.9	Spleen Pool	7.6
Breast Pool	29.5	Thymus Pool	11.0
Trachea	6.3	CNS cancer (glio/astro) U87-MG	45.4
Lung	6.6	CNS cancer (glio/astro) U-118-MG	38.7
Fetal Lung	24.8	CNS cancer (neuro;met) SK-N-AS	9.8
Lung ca. NCI-N417	4.2	CNS cancer (astro) SF-539	14.7
Lung ca. LX-1	31.0	CNS cancer (astro) SNB-75	100.0
Lung ca. NCI-H146	8.5	CNS cancer (glio) SNB-19	21.5
Lung ca. SHP-77	16.2	CNS cancer (glio) SF-295	71.7
Lung ca. A549	25.0	Brain (Amygdala) Pool	7.6
Lung ca. NCI-H526	5.5	Brain (cerebellum)	23.3
Lung ca. NCI-H23	34.2	Brain (fetal)	26.4
Lung ca. NCI-H460	37.6	Brain (Hippocampus) Pool	8.0
Lung ca. HOP-62	13.5	Cerebral Cortex Pool	8.4
Lung ca. NCI-H522	20.2	Brain (Substantia nigra) Pool	7.6
Liver	0.2	Brain (Thalamus) Pool	14.0
Fetal Liver	31.6	Brain (whole)	6.7
Liver ca. HepG2	16.0	Spinal Cord Pool	9.9
Kidney Pool	32.8	Adrenal Gland	5.3
Fetal Kidney	11.7	Pituitary gland Pool	2.9

Renal ca. 786-0	11.0	Salivary Gland	1.6
Renal ca. A498	6.2	Thyroid (female)	7.7
Renal ca. ACHN	7.6	Pancreatic ca. CAPAN2	12.2
Renal ca. UO-31	16.8	Pancreas Pool	18.7

Table PD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4961, Run 223691546	Tissue Name	Rel. Exp.(%) Ag4961, Run 223691546
Secondary Th1 act	63.7	HUVEC IL-1beta	48.6
Secondary Th2 act	46.7	HUVEC IFN gamma	45.7
Secondary Tr1 act	54.3	HUVEC TNF alpha + IFN gamma	28.1
Secondary Th1 rest	7.7	HUVEC TNF alpha + IL4	30.6
Secondary Th2 rest	16.8	HUVEC IL-11	15.0
Secondary Tr1 rest	22.2	Lung Microvascular EC none	0.0
Primary Th1 act	27.5	Lung Microvascular EC TNFalpha + IL-1beta	49.3
Primary Th2 act	37.1	Microvascular Dermal EC none	38.4
Primary Tr1 act	39.2	Microvascular Dermal EC TNFalpha + IL-1beta	32.8
Primary Th1 rest	6.6	Bronchial epithelium TNFalpha + IL1beta	30.8
Primary Th2 rest	4.8	Small airway epithelium none	14.4
Primary Tr1 rest	9.2	Small airway epithelium TNFalpha + IL-1beta	31.0
CD45RA CD4 lymphocyte act	28.7	Coronary artery SMC rest	29.3
CD45RO CD4 lymphocyte act	25.2	Coronary artery SMC TNFalpha + IL-1beta	34.6
CD8 lymphocyte act	43.2	Astrocytes rest	29.5
Secondary CD8 lymphocyte rest	62.0	Astrocytes TNFalpha + IL-1beta	27.5
Secondary CD8 lymphocyte act	23.2	KU-812 (Basophil) rest	48.3
CD4 lymphocyte none	5.3	KU-812 (Basophil) PMA/ionomycin	100.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	16.0	CCD1106 (Keratinocytes) none	31.4
LAK cells rest	32.1	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	16.6
LAK cells IL-2	30.6	Liver cirrhosis	6.8
LAK cells IL-2+IL-12	14.5	NCI-H292 none	33.0
LAK cells IL-2+IFN gamma	22.2	NCI-H292 IL-4	47.3
LAK cells IL-2+ IL-18	29.3	NCI-H292 IL-9	68.8
LAK cells PMA/ionomycin	47.6	NCI-H292 IL-13	44.1
NK Cells IL-2 rest	30.1	NCI-H292 IFN gamma	32.1

Two Way MLR 3 day	32.3	HPAEC none	25.3
Two Way MLR 5 day	18.6	HPAEC TNF alpha + IL-1 beta	44.8
Two Way MLR 7 day	17.9	Lung fibroblast none	30.6
PBMC rest	4.9	Lung fibroblast TNF alpha + IL-1 beta	18.7
PBMC PWM	17.0	Lung fibroblast IL-4	24.1
PBMC PHA-L	18.4	Lung fibroblast IL-9	29.9
Ramos (B cell) none	44.8	Lung fibroblast IL-13	52.9
Ramos (B cell) ionomycin	58.6	Lung fibroblast IFN gamma	57.0
B lymphocytes PWM	27.9	Dermal fibroblast CCD1070 rest	58.6
B lymphocytes CD40L and IL-4	26.4	Dermal fibroblast CCD1070 TNF alpha	62.9
EOL-1 dbcAMP	36.3	Dermal fibroblast CCD1070 IL-1 beta	31.0
EOL-1 dbcAMP PMA/ionomycin	56.3	Dermal fibroblast IFN gamma	16.5
Dendritic cells none	25.9	Dermal fibroblast IL-4	28.1
Dendritic cells LPS	32.8	Dermal Fibroblasts rest	17.0
Dendritic cells anti-CD40	25.9	Neutrophils TNFa+LPS	7.2
Monocytes rest	25.7	Neutrophils rest	10.2
Monocytes LPS	50.3	Colon	9.6
Macrophages rest	34.4	Lung	15.4
Macrophages LPS	13.2	Thymus	24.5
HUVEC none	33.2	Kidney	19.8
HUVEC starved	25.9		

CNS_neurodegeneration_v1.0 Summary: Ag4961 This panel confirms the expression of this gene at low levels in the brains of an independent group of individuals. However, no differential expression of this gene was detected between Alzheimer's diseased postmortem brains and those of non-demented controls in this experiment. Please see Panel 1.5 for a discussion of the potential role of this gene in treatment of central nervous system disorders.

General_screening_panel_v1.5 Summary: Ag4961 Highest expression of this gene is detected in brain cancer SNB-75 cell line (CT=28.6). Moderate levels of expression of this gene is also seen in cluster of cancer cell lines derived from pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers. Thus, expression of this gene could be used as a marker to detect the presence of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of pancreatic, gastric, colon, lung, liver, renal, breast, ovarian, prostate, squamous cell carcinoma, melanoma and brain cancers.

Among tissues with metabolic or endocrine function, this gene is expressed at moderate to low levels in pancreas, adipose, adrenal gland, thyroid, pituitary gland, skeletal muscle, heart, fetal liver and the gastrointestinal tract. Therefore, therapeutic modulation of the activity of this gene may prove useful in the treatment of endocrine/metabolically related diseases, such as obesity and diabetes.

Interestingly, this gene is expressed at much higher levels in fetal (CT=30.2) when compared to adult liver (CT=37). This observation suggests that expression of this gene can be used to distinguish fetal from adult liver. In addition, the relative overexpression of this gene in fetal tissue suggests that the protein product may enhance liver growth or development in the fetus and thus may also act in a regenerative capacity in the adult. Therefore, therapeutic modulation of the protein encoded by this gene could be useful in treatment of liver related diseases.

In addition, this gene is expressed at moderate levels in all regions of the central nervous system examined, including amygdala, hippocampus, substantia nigra, thalamus, cerebellum, cerebral cortex, and spinal cord. Therefore, therapeutic modulation of this gene product may be useful in the treatment of central nervous system disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

Panel 4.1D Summary: Ag4961 Highest expression of this gene is detected in PMA/ionomycin stimulated basophil (CT=30.4). This gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

Q. NOV72a: WW domain containing protein

Expression of gene NOV72a was assessed using the primer-probe set Ag4977, described in Table QA. Results of the RTQ-PCR runs are shown in Tables QB and QC.

Table QA. Probe Name Ag4977

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-taggatgggaagaggcatatg-3'	21	263	396
Probe	TET-5'-cttcataagaccacacaccaaacca-3'-TAMRA	26	303	397
Reverse	5'-tactcgaggatcctcaatctga-3'	22	330	398

Table QB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4977, Run 228940919	Tissue Name	Rel. Exp.(%) Ag4977, Run 228940919
Adipose	0.9	Renal ca. TK-10	40.6
Melanoma* Hs688(A).T	0.1	Bladder	15.0
Melanoma* Hs688(B).T	0.2	Gastric ca. (liver met.) NCI-N87	23.5
Melanoma* M14	3.5	Gastric ca. KATO III	59.5
Melanoma* LOXIMVI	6.3	Colon ca. SW-948	7.4
Melanoma* SK-MEL-5	8.4	Colon ca. SW480	29.9
Squamous cell carcinoma SCC-4	9.1	Colon ca.* (SW480 met) SW620	5.8
Testis Pool	0.5	Colon ca. HT29	8.4
Prostate ca.* (bone met) PC-3	2.3	Colon ca. HCT-116	21.3
Prostate Pool	4.1	Colon ca. CaCo-2	15.4
Placenta	8.7	Colon cancer tissue	4.6
Uterus Pool	0.2	Colon ca. SW1116	1.5
Ovarian ca. OVCAR-3	12.8	Colon ca. Colo-205	3.6
Ovarian ca. SK-OV-3	31.2	Colon ca. SW-48	2.4
Ovarian ca. OVCAR-4	66.0	Colon Pool	0.3
Ovarian ca. OVCAR-5	100.0	Small Intestine Pool	0.4
Ovarian ca. IGROV-1	18.3	Stomach Pool	1.1
Ovarian ca. OVCAR-8	6.2	Bone Marrow Pool	0.3
Ovary	1.9	Fetal Heart	0.0
Breast ca. MCF-7	7.0	Heart Pool	0.0
Breast ca. MDA-MB-231	18.0	Lymph Node Pool	0.6
Breast ca. BT 549	2.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	6.3	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	0.1
Breast Pool	1.6	Thymus Pool	1.5

Trachea	4.7	CNS cancer (glio/astro) U87-MG	0.7
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	4.9	CNS cancer (neuro;met) SK-N-AS	3.1
Lung ca. NCI-N417	0.3	CNS cancer (astro) SF-539	0.3
Lung ca. LX-1	6.7	CNS cancer (astro) SNB-75	0.9
Lung ca. NCI-H146	2.4	CNS cancer (glio) SNB-19	18.6
Lung ca. SHP-77	0.9	CNS cancer (glio) SF-295	2.2
Lung ca. A549	16.7	Brain (Amygdala) Pool	3.8
Lung ca. NCI-H526	2.0	Brain (cerebellum)	8.0
Lung ca. NCI-H23	3.2	Brain (fetal)	15.9
Lung ca. NCI-H460	1.6	Brain (Hippocampus) Pool	5.0
Lung ca. HOP-62	1.8	Cerebral Cortex Pool	3.9
Lung ca. NCI-H522	6.6	Brain (Substantia nigra) Pool	4.2
Liver	0.5	Brain (Thalamus) Pool	6.3
Fetal Liver	2.5	Brain (whole)	4.8
Liver ca. HepG2	4.6	Spinal Cord Pool	5.0
Kidney Pool	0.1	Adrenal Gland	0.4
Fetal Kidney	5.2	Pituitary gland Pool	0.4
Renal ca. 786-0	9.8	Salivary Gland	21.0
Renal ca. A498	7.9	Thyroid (female)	3.9
Renal ca. ACHN	23.0	Pancreatic ca. CAPAN2	33.2
Renal ca. UO-31	34.2	Pancreas Pool	5.1

Table QC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4977, Run 223692679	Tissue Name	Rel. Exp.(%) Ag4977, Run 223692679
Secondary Th1 act	0.1	HUVEC IL-1beta	0.3
Secondary Th2 act	0.0	HUVEC IFN gamma	0.2
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	0.0	HUVEC TNF alpha + IL4	0.5
Secondary Th2 rest	0.0	HUVEC IL-11	0.1
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.9	Lung Microvascular EC TNFalpha + IL-1beta	1.3
Primary Th2 act	0.1	Microvascular Dermal EC none	0.0
Primary Tr1 act	0.7	Microvascular Dermal EC TNFalpha + IL-1beta	0.3
Primary Th1 rest	0.0	Bronchial epithelium TNFalpha + IL1beta	35.1
Primary Th2 rest	0.0	Small airway epithelium none	24.5

Primary Tr1 rest	0.0	Small airway epithelium TNFalpha + IL-1beta	55.9
CD45RA CD4 lymphocyte act	12.6	Coronary artery SMC rest	6.7
CD45RO CD4 lymphocyte act	1.1	Coronary artery SMC TNFalpha + IL-1beta	10.7
CD8 lymphocyte act	0.0	Astrocytes rest	14.6
Secondary CD8 lymphocyte rest	1.7	Astrocytes TNFalpha + IL-1beta	60.7
Secondary CD8 lymphocyte act	0.0	KU-812 (Basophil) rest	3.2
CD4 lymphocyte none	0.1	KU-812 (Basophil) PMA/ionomycin	5.7
2ry Th1/Th2/Tr1_anti-CD95 CH11	0.0	CCD1106 (Keratinocytes) none	28.9
LAK cells rest	0.0	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	57.4
LAK cells IL-2	0.0	Liver cirrhosis	9.9
LAK cells IL-2+IL-12	0.0	NCI-H292 none	40.6
LAK cells IL-2+IFN gamma	0.0	NCI-H292 IL-4	68.8
LAK cells IL-2+ IL-18	0.3	NCI-H292 IL-9	100.0
LAK cells PMA/ionomycin	0.0	NCI-H292 IL-13	70.7
NK Cells IL-2 rest	0.1	NCI-H292 IFN gamma	82.4
Two Way MLR 3 day	0.0	HPAEC none	0.0
Two Way MLR 5 day	0.1	HPAEC TNF alpha + IL-1 beta	1.5
Two Way MLR 7 day	0.0	Lung fibroblast none	4.2
PBMC rest	0.1	Lung fibroblast TNF alpha + IL-1 beta	8.2
PBMC PWM	0.0	Lung fibroblast IL-4	4.4
PBMC PHA-L	0.2	Lung fibroblast IL-9	8.9
Ramos (B cell) none	31.4	Lung fibroblast IL-13	3.7
Ramos (B cell) ionomycin	23.2	Lung fibroblast IFN gamma	4.7
B lymphocytes PWM	1.0	Dermal fibroblast CCD1070 rest	5.4
B lymphocytes CD40L and IL-4	1.2	Dermal fibroblast CCD1070 TNF alpha	23.0
EOL-1 dbcAMP	3.1	Dermal fibroblast CCD1070 IL-1 beta	30.4
EOL-1 dbcAMP PMA/ionomycin	0.2	Dermal fibroblast IFN gamma	2.5
Dendritic cells none	0.0	Dermal fibroblast IL-4	2.8
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	3.7
Dendritic cells anti-CD40	0.0	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	9.9
Macrophages rest	0.4	Lung	12.7
Macrophages LPS	0.1	Thymus	1.1
HUVEC none	0.0	Kidney	79.6
HUVEC starved	0.1		

General_screening_panel_v1.5 Summary: Ag4977 Highest expression of this gene is seen in an ovarian cancer cell line (CT=25.8). This gene is widely expressed in this panel,

with high to moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer, specifically in ovarian, colon and renal cancers.

5 In addition, this gene is expressed at much higher levels in fetal lung tissue (CT=30) when compared to expression in the adult counterpart (CT=37). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue.

Among tissues with metabolic function, this gene is expressed at moderate to low levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal liver. This
10 widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the
15 hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag4977 Highest expression of this gene is seen in IL-9
20 treated NCI-H292 cells (CT=28.5). Expression of this gene is also seen at moderate levels in activated and resting keratinocytes, small airway epithelium, bronchial epithelium, and astrocytes, a cluster of samples derived from NCI-H292 cells, and normal kidney. Low but significant levels of expression are seen in treated and untreated dermal and lung fibroblasts. The expression of this gene in cells derived from the lung and skin suggests that this gene
25 may be involved in normal conditions as well as pathological and inflammatory lung and skin disorders that include chronic obstructive pulmonary disease, asthma, allergy, psoriasis and emphysema.

R. NOV73a and NOV73b: GASDERMIN

Expression of gene NOV73a and full length physical clone NOV73b was assessed
30 using the primer-probe set Ag4981, described in Table RA. Results of the RTQ-PCR runs are shown in Tables RB and RC. Please note that NOV73b represents a full-length physical clone of the NOV73a gene, validating the prediction of the gene sequence.

Table RA. Probe Name Ag4981

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aggcactcccaaaagatgc-3'	20	1007	399
Probe	TET-5'-ccatcctctatttcgttgaggcccta-3'-TAMRA	26	1052	400
Reverse	5'-agcttcgttgggttcact-3'	20	1087	401

Table RB. General screening panel v1.5

5

Tissue Name	Rel. Exp.(%) Ag4981, Run 228940922	Tissue Name	Rel. Exp.(%) Ag4981, Run 228940922
Adipose	2.1	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	0.4	Bladder	1.8
Melanoma* Hs688(B).T	1.4	Gastric ca. (liver met.) NCI-N87	3.1
Melanoma* M14	0.0	Gastric ca. KATO III	0.1
Melanoma* LOXIMVI	6.9	Colon ca. SW-948	0.9
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	10.6
Squamous cell carcinoma SCC-4	0.7	Colon ca.* (SW480 met) SW620	1.9
Testis Pool	3.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	0.0	Colon ca. HCT-116	0.0
Prostate Pool	0.0	Colon ca. CaCo-2	0.1
Placenta	0.9	Colon cancer tissue	100.0
Uterus Pool	0.7	Colon ca. SW1116	0.1
Ovarian ca. OVCAR-3	0.0	Colon ca. Colo-205	34.4
Ovarian ca. SK-OV-3	0.7	Colon ca. SW-48	31.6
Ovarian ca. OVCAR-4	0.1	Colon Pool	2.0
Ovarian ca. OVCAR-5	1.2	Small Intestine Pool	0.3
Ovarian ca. IGROV-1	0.0	Stomach Pool	0.7
Ovarian ca. OVCAR-8	0.4	Bone Marrow Pool	0.7
Ovary	0.7	Fetal Heart	0.0
Breast ca. MCF-7	0.0	Heart Pool	0.7
Breast ca. MDA-MB-231	0.0	Lymph Node Pool	1.5
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.6
Breast ca. T47D	0.0	Skeletal Muscle Pool	0.5
Breast ca. MDA-N	0.2	Spleen Pool	0.0
Breast Pool	0.6	Thymus Pool	0.6
Trachea	1.8	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.5	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	6.3	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0

Lung ca. LX-1	82.4	CNS cancer (astro) SNB-75	0.8
Lung ca. NCI-H146	0.8	CNS cancer (glio) SNB-19	0.0
Lung ca. SHP-77	1.1	CNS cancer (glio) SF-295	0.0
Lung ca. A549	0.0	Brain (Amygdala) Pool	0.3
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	0.0	Brain (fetal)	0.1
Lung ca. NCI-H460	0.0	Brain (Hippocampus) Pool	0.2
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	0.0
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	0.0
Liver	0.0	Brain (Thalamus) Pool	0.2
Fetal Liver	1.8	Brain (whole)	0.4
Liver ca. HepG2	0.0	Spinal Cord Pool	0.5
Kidney Pool	1.8	Adrenal Gland	1.9
Fetal Kidney	12.3	Pituitary gland Pool	0.4
Renal ca. 786-0	0.9	Salivary Gland	1.0
Renal ca. A498	0.2	Thyroid (female)	0.1
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	0.9
Renal ca. UO-31	0.2	Pancreas Pool	3.0

Table RC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4981, Run 223693389	Tissue Name	Rel. Exp.(%) Ag4981, Run 223693389
Secondary Th1 act	3.6	HUVEC IL-1beta	0.0
Secondary Th2 act	6.3	HUVEC IFN gamma	0.0
Secondary Tr1 act	0.0	HUVEC TNF alpha + IFN gamma	0.0
Secondary Th1 rest	3.0	HUVEC TNF alpha + IL4	0.0
Secondary Th2 rest	16.6	HUVEC IL-11	0.0
Secondary Tr1 rest	0.0	Lung Microvascular EC none	0.0
Primary Th1 act	0.0	Lung Microvascular EC TNFalpha + IL-1beta	1.7
Primary Th2 act	29.1	Microvascular Dermal EC none	0.0
Primary Tr1 act	14.6	Microvascular Dermal EC TNFalpha + IL-1beta	0.0
Primary Th1 rest	9.3	Bronchial epithelium TNFalpha + IL1beta	8.4
Primary Th2 rest	9.9	Small airway epithelium none	100.0
Primary Tr1 rest	62.4	Small airway epithelium TNFalpha + IL-1beta	44.8
CD45RA CD4 lymphocyte act	12.9	Coronary artery SMC rest	0.0
CD45RO CD4 lymphocyte act	84.7	Coronary artery SMC TNFalpha + IL-1beta	6.8

CD8 lymphocyte act	24.5	Astrocytes rest	1.4
Secondary CD8 lymphocyte rest	29.3	Astrocytes TNFalpha + IL-1 beta	0.0
Secondary CD8 lymphocyte act	2.0	KU-812 (Basophil) rest	0.0
CD4 lymphocyte none	12.9	KU-812 (Basophil) PMA/ionomycin	0.0
2ry Th1/Th2/Tr1_anti-CD95 CH11	14.7	CCD1106 (Keratinocytes) none	4.2
LAK cells rest	5.2	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	29.1
LAK cells IL-2	16.6	Liver cirrhosis	0.0
LAK cells IL-2+IL-12	3.4	NCI-H292 none	0.0
LAK cells IL-2+IFN gamma	4.4	NCI-H292 IL-4	0.0
LAK cells IL-2+ IL-18	0.0	NCI-H292 IL-9	0.0
LAK cells PMA/ionomycin	1.1	NCI-H292 IL-13	0.0
NK Cells IL-2 rest	7.3	NCI-H292 IFN gamma	0.0
Two Way MLR 3 day	8.7	HPAEC none	0.0
Two Way MLR 5 day	0.0	HPAEC TNF alpha + IL-1 beta	0.0
Two Way MLR 7 day	2.2	Lung fibroblast none	0.0
PBMC rest	14.2	Lung fibroblast TNF alpha + IL-1 beta	0.0
PBMC PWM	4.5	Lung fibroblast IL-4	0.0
PBMC PHA-L	14.0	Lung fibroblast IL-9	0.5
Ramos (B cell) none	2.6	Lung fibroblast IL-13	0.0
Ramos (B cell) ionomycin	4.0	Lung fibroblast IFN gamma	0.0
B lymphocytes PWM	0.0	Dermal fibroblast CCD1070 rest	4.3
B lymphocytes CD40L and IL-4	0.0	Dermal fibroblast CCD1070 TNF alpha	7.1
EOL-1 dbcAMP	0.0	Dermal fibroblast CCD1070 IL-1 beta	6.3
EOL-1 dbcAMP PMA/ionomycin	0.0	Dermal fibroblast IFN gamma	0.0
Dendritic cells none	10.7	Dermal fibroblast IL-4	0.0
Dendritic cells LPS	0.0	Dermal Fibroblasts rest	0.0
Dendritic cells anti-CD40	2.2	Neutrophils TNFa+LPS	0.0
Monocytes rest	0.0	Neutrophils rest	0.0
Monocytes LPS	0.0	Colon	3.6
Macrophages rest	29.7	Lung	11.1
Macrophages LPS	0.0	Thymus	7.4
HUVEC none	0.0	Kidney	2.1
HUVEC starved	0.0		

General_screening_panel_v1.5 Summary: Ag4981 Highest expression is seen in colon cancer tissue (CT=28.4), with prominent expression also detected in a lung cancer cell line. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker to detect the presence of colon cancer.

Furthermore, therapeutic modulation of the expression or function of this gene may be effective in the treatment of colon cancer.

Low but significant levels of expression are seen in pancreas, metabolic, and adrenal suggesting that modulation of the expression or function of this gene may be useful in the treatment of metabolic disease, including obesity and diabetes.

Panel 4.1D Summary: Ag4981 Highest expression of this gene is seen in small airway epithelium (CT=31.2). Low but significant levels of expression are detected in TNF- α and IL-1 β treated keratinocytes and small airway epithelium, primary T cells, dendritic cells, macrophages and normal lung.

10 S. NOV75a: Synaptotagmin-Like Protein 3-A

Expression of gene NOV75a was assessed using the primer-probe set Ag4993, described in Table SA. Results of the RTQ-PCR runs are shown in Tables SB and SC.

Table SA. Probe Name Ag4993

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aaagtgcattccgtatgtgaag-3'	22	1019	402
Probe	TET-5'-acctacctgtgcccgcagatcct-3'-TAMRA	25	1041	403
Reverse	5'-gtgttcctttggactccagtct-3'	22	1081	404

Table SB. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag4993, Run 228941089	Tissue Name	Rel. Exp.(%) Ag4993, Run 228941089
Adipose	1.5	Renal ca. TK-10	2.6
Melanoma* Hs688(A).T	1.2	Bladder	2.7
Melanoma* Hs688(B).T	2.6	Gastric ca. (liver met.) NCI-N87	13.7
Melanoma* M14	0.4	Gastric ca. KATO III	0.2
Melanoma* LOXIMVI	7.7	Colon ca. SW-948	0.1
Melanoma* SK-MEL-5	0.4	Colon ca. SW480	3.6
Squamous cell carcinoma SCC-4	8.9	Colon ca.* (SW480 met) SW620	0.9
Testis Pool	1.7	Colon ca. HT29	0.2
Prostate ca.* (bone met) PC-3	1.0	Colon ca. HCT-116	100.0
Prostate Pool	2.7	Colon ca. CaCo-2	0.8
Placenta	0.4	Colon cancer tissue	1.3
Uterus Pool	0.9	Colon ca. SW1116	0.2
Ovarian ca. OVCAR-3	1.5	Colon ca. Colo-205	0.4

Ovarian ca. SK-OV-3	3.8	Colon ca. SW-48	0.2
Ovarian ca. OVCAR-4	0.0	Colon Pool	1.4
Ovarian ca. OVCAR-5	5.8	Small Intestine Pool	1.2
Ovarian ca. IGROV-1	0.8	Stomach Pool	1.2
Ovarian ca. OVCAR-8	2.1	Bone Marrow Pool	0.8
Ovary	0.8	Fetal Heart	3.0
Breast ca. MCF-7	0.4	Heart Pool	0.6
Breast ca. MDA-MB-231	0.8	Lymph Node Pool	1.5
Breast ca. BT 549	1.6	Fetal Skeletal Muscle	2.1
Breast ca. T47D	1.2	Skeletal Muscle Pool	6.2
Breast ca. MDA-N	0.2	Spleen Pool	3.7
Breast Pool	1.2	Thymus Pool	9.2
Trachea	4.4	CNS cancer (glio/astro) U87-MG	18.7
Lung	0.4	CNS cancer (glio/astro) U-118-MG	12.5
Fetal Lung	5.8	CNS cancer (neuro;met) SK-N-AS	1.2
Lung ca. NCI-N417	0.1	CNS cancer (astro) SF-539	1.1
Lung ca. LX-1	4.6	CNS cancer (astro) SNB-75	1.6
Lung ca. NCI-H146	4.9	CNS cancer (glio) SNB-19	0.9
Lung ca. SHP-77	4.9	CNS cancer (glio) SF-295	1.5
Lung ca. A549	0.5	Brain (Amygdala) Pool	0.7
Lung ca. NCI-H526	0.1	Brain (cerebellum)	0.8
Lung ca. NCI-H23	1.0	Brain (fetal)	1.4
Lung ca. NCI-H460	0.3	Brain (Hippocampus) Pool	1.4
Lung ca. HOP-62	0.5	Cerebral Cortex Pool	0.4
Lung ca. NCI-H522	1.0	Brain (Substantia nigra) Pool	0.7
Liver	0.1	Brain (Thalamus) Pool	1.2
Fetal Liver	0.1	Brain (whole)	0.8
Liver ca. HepG2	0.2	Spinal Cord Pool	1.4
Kidney Pool	2.7	Adrenal Gland	1.0
Fetal Kidney	1.1	Pituitary gland Pool	0.2
Renal ca. 786-0	47.0	Salivary Gland	3.7
Renal ca. A498	7.9	Thyroid (female)	9.9
Renal ca. ACHN	1.9	Pancreatic ca. CAPAN2	2.2
Renal ca. UO-31	2.0	Pancreas Pool	1.9

Table SC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag4993, Run 223739949	Tissue Name	Rel. Exp.(%) Ag4993, Run 223739949
Secondary Th1 act	88.3	HUVEC IL-1beta	0.6
Secondary Th2 act	100.0	HUVEC IFN gamma	3.6

Secondary Tr1 act	62.4	HUVEC TNF alpha + IFN gamma	4.7
Secondary Th1 rest	11.4	HUVEC TNF alpha + IL4	0.2
Secondary Th2 rest	17.2	HUVEC IL-11	0.5
Secondary Tr1 rest	10.7	Lung Microvascular EC none	3.7
Primary Th1 act	13.0	Lung Microvascular EC TNFalpha + IL-1beta	4.9
Primary Th2 act	55.9	Microvascular Dermal EC none	2.0
Primary Tr1 act	27.0	Microvascular Dermal EC TNFalpha + IL-1beta	5.0
Primary Th1 rest	11.8	Bronchial epithelium TNFalpha + IL1beta	1.6
Primary Th2 rest	10.2	Small airway epithelium none	0.3
Primary Tr1 rest	15.3	Small airway epithelium TNFalpha + IL-1beta	0.7
CD45RA CD4 lymphocyte act	29.9	Coronary artery SMC rest	4.2
CD45RO CD4 lymphocyte act	52.5	Coronary artery SMC TNFalpha + IL-1beta	5.5
CD8 lymphocyte act	37.4	Astrocytes rest	0.2
Secondary CD8 lymphocyte rest	71.7	Astrocytes TNFalpha + IL-1beta	0.6
Secondary CD8 lymphocyte act	34.6	KU-812 (Basophil) rest	1.0
CD4 lymphocyte none	8.2	KU-812 (Basophil) PMA/ionomycin	1.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	30.1	CCD1106 (Keratinocytes) none	3.7
LAK cells rest	11.7	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	2.1
LAK cells IL-2	26.4	Liver cirrhosis	4.2
LAK cells IL-2+IL-12	39.2	NCI-H292 none	13.8
LAK cells IL-2+IFN gamma	25.7	NCI-H292 IL-4	13.7
LAK cells IL-2+ IL-18	46.3	NCI-H292 IL-9	23.3
LAK cells PMA/ionomycin	42.3	NCI-H292 IL-13	13.4
NK Cells IL-2 rest	44.8	NCI-H292 IFN gamma	21.0
Two Way MLR 3 day	15.7	HPAEC none	0.5
Two Way MLR 5 day	21.6	HPAEC TNF alpha + IL-1 beta	4.0
Two Way MLR 7 day	23.5	Lung fibroblast none	1.9
PBMC rest	6.9	Lung fibroblast TNF alpha + IL-1 beta	0.9
PBMC PWM	77.4	Lung fibroblast IL-4	0.3
PBMC PHA-L	56.6	Lung fibroblast IL-9	0.6
Ramos (B cell) none	1.3	Lung fibroblast IL-13	0.4
Ramos (B cell) ionomycin	2.0	Lung fibroblast IFN gamma	1.7
B lymphocytes PWM	22.4	Dermal fibroblast CCD1070 rest	1.2
B lymphocytes CD40L and IL-4	5.2	Dermal fibroblast CCD1070 TNF alpha	18.0
EOL-1 dbcAMP	6.2	Dermal fibroblast CCD1070 IL-1 beta	0.9
EOL-1 dbcAMP PMA/ionomycin	17.7	Dermal fibroblast IFN gamma	2.7
Dendritic cells none	4.7	Dermal fibroblast IL-4	1.2

Dendritic cells LPS	4.1	Dermal Fibroblasts rest	1.6
Dendritic cells anti-CD40	2.1	Neutrophils TNFa+LPS	1.8
Monocytes rest	14.7	Neutrophils rest	13.5
Monocytes LPS	6.8	Colon	3.3
Macrophages rest	7.5	Lung	3.0
Macrophages LPS	7.5	Thymus	22.7
HUVEC none	0.5	Kidney	1.4
HUVEC starved	0.5		

General_screening_panel_v1.5 Summary: Ag4993 Highest expression of this gene is seen in a colon cancer cell line (CT=25.3). High levels of expression are also seen in a renal cancer cell line. Thus, expression of this gene could be used to differentiate between these samples and other samples on this panel and as a marker of these cancers. In addition, therapeutic modulation of the expression or function of this gene may be useful in the treatment of these cancers.

In addition, this gene is expressed at much higher levels in fetal lung tissue (CT=29.4) when compared to expression in the adult counterpart (CT=33.3). Thus, expression of this gene may be used to differentiate between the fetal and adult source of this tissue. In addition, modulation of the expression or function of this gene may be useful in the treatment of diseases that affect this tissue.

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag4993 Highest expression of this gene is seen in chronically activated Th2 cells (Ct=26.2). This gene is widely expressed on this panel, but is more prominently expressed in T cells, B cells and LAK cells. Therefore, the putative protein

encoded by this gene could potentially be used diagnostically to identify B or T cells. In addition, the gene product could also potentially be used therapeutically in the treatment of asthma, emphysema, IBD, lupus or arthritis and in other diseases in which T cells and B cells are involved.

5 T. NOV76a: Hypothetical Intracellular

Expression of gene NOV76a was assessed using the primer-probe set Ag5023, described in Table TA. Results of the RTQ-PCR runs are shown in Tables TB, TC and TD.

Table TA. Probe Name Ag5023

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cgtgtcaacggtatgtttctct-3'	22	1175	405
Probe	TET-5'-taattcttgacatccaaggatggca-3'-TAMRA	26	1221	406
Reverse	5'-agagaaagcaatgggtccattt-3'	22	1249	407

10

Table TB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag5023, Run 224757456	Tissue Name	Rel. Exp.(%) Ag5023, Run 224757456
AD 1 Hippo	19.6	Control (Path) 3 Temporal Ctx	7.1
AD 2 Hippo	25.9	Control (Path) 4 Temporal Ctx	18.3
AD 3 Hippo	9.7	AD 1 Occipital Ctx	20.2
AD 4 Hippo	7.1	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	73.7	AD 3 Occipital Ctx	12.9
AD 6 Hippo	71.7	AD 4 Occipital Ctx	18.8
Control 2 Hippo	21.0	AD 5 Occipital Ctx	20.0
Control 4 Hippo	9.6	AD 6 Occipital Ctx	31.2
Control (Path) 3 Hippo	4.7	Control 1 Occipital Ctx	4.2
AD 1 Temporal Ctx	26.8	Control 2 Occipital Ctx	40.1
AD 2 Temporal Ctx	37.9	Control 3 Occipital Ctx	12.4
AD 3 Temporal Ctx	6.8	Control 4 Occipital Ctx	5.9
AD 4 Temporal Ctx	15.1	Control (Path) 1 Occipital Ctx	69.7
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	9.6
AD 5 Sup Temporal Ctx	39.0	Control (Path) 3 Occipital Ctx	4.7
AD 6 Inf Temporal Ctx	56.3	Control (Path) 4 Occipital Ctx	23.0
AD 6 Sup Temporal Ctx	61.1	Control 1 Parietal Ctx	9.1
Control 1 Temporal Ctx	14.6	Control 2 Parietal Ctx	52.9
Control 2 Temporal Ctx	25.9	Control 3 Parietal Ctx	17.8
Control 3 Temporal Ctx	18.2	Control (Path) 1 Parietal Ctx	55.1

Control 4 Temporal Ctx	4.2	Control (Path) 2 Parietal Ctx	17.4
Control (Path) 1 Temporal Ctx	35.6	Control (Path) 3 Parietal Ctx	5.7
Control (Path) 2 Temporal Ctx	29.3	Control (Path) 4 Parietal Ctx	38.7

Table TC. General screening panel v1.5

Tissue Name	Rel. Exp.(%) Ag5023, Run 228959376	Tissue Name	Rel. Exp.(%) Ag5023, Run 228959376
Adipose	2.3	Renal ca. TK-10	5.7
Melanoma* Hs688(A).T	2.3	Bladder	3.6
Melanoma* Hs688(B).T	3.0	Gastric ca. (liver met.) NCI-N87	7.9
Melanoma* M14	2.1	Gastric ca. KATO III	8.4
Melanoma* LOXIMVI	2.7	Colon ca. SW-948	2.1
Melanoma* SK-MEL-5	3.3	Colon ca. SW480	1.9
Squamous cell carcinoma SCC-4	2.5	Colon ca. * (SW480 met) SW620	1.3
Testis Pool	2.0	Colon ca. HT29	1.3
Prostate ca. * (bone met) PC-3	4.2	Colon ca. HCT-116	5.3
Prostate Pool	2.5	Colon ca. CaCo-2	3.3
Placenta	0.7	Colon cancer tissue	1.5
Uterus Pool	2.4	Colon ca. SW1116	0.5
Ovarian ca. OVCAR-3	4.5	Colon ca. Colo-205	1.1
Ovarian ca. SK-OV-3	11.1	Colon ca. SW-48	0.5
Ovarian ca. OVCAR-4	1.3	Colon Pool	5.3
Ovarian ca. OVCAR-5	5.2	Small Intestine Pool	5.0
Ovarian ca. IGROV-1	3.2	Stomach Pool	3.0
Ovarian ca. OVCAR-8	2.4	Bone Marrow Pool	1.5
Ovary	3.0	Fetal Heart	3.1
Breast ca. MCF-7	4.5	Heart Pool	2.5
Breast ca. MDA-MB-231	9.9	Lymph Node Pool	4.9
Breast ca. BT 549	5.1	Fetal Skeletal Muscle	4.5
Breast ca. T47D	1.4	Skeletal Muscle Pool	10.7
Breast ca. MDA-N	1.8	Spleen Pool	2.1
Breast Pool	4.1	Thymus Pool	2.5
Trachea	1.4	CNS cancer (glio/astro) U87-MG	16.6
Lung	0.0	CNS cancer (glio/astro) U-118-MG	11.2
Fetal Lung	8.0	CNS cancer (neuro;met) SK-N-AS	7.1
Lung ca. NCI-N417	1.1	CNS cancer (astro) SF-539	2.3
Lung ca. LX-1	4.2	CNS cancer (astro) SNB-75	5.9
Lung ca. NCI-H146	0.6	CNS cancer (glio) SNB-19	3.6
Lung ca. SHP-77	2.1	CNS cancer (glio) SF-295	12.1
Lung ca. A549	1.6	Brain (Amygdala) Pool	1.3

Lung ca. NCI-H526	1.0	Brain (cerebellum)	4.4
Lung ca. NCI-H23	2.1	Brain (fetal)	2.1
Lung ca. NCI-H460	0.6	Brain (Hippocampus) Pool	0.7
Lung ca. HOP-62	1.3	Cerebral Cortex Pool	0.6
Lung ca. NCI-H522	1.1	Brain (Substantia nigra) Pool	0.6
Liver	0.0	Brain (Thalamus) Pool	0.9
Fetal Liver	100.0	Brain (whole)	0.2
Liver ca. HepG2	2.2	Spinal Cord Pool	2.2
Kidney Pool	7.4	Adrenal Gland	1.8
Fetal Kidney	4.6	Pituitary gland Pool	1.0
Renal ca. 786-0	6.2	Salivary Gland	0.9
Renal ca. A498	0.7	Thyroid (female)	1.4
Renal ca. ACHN	2.4	Pancreatic ca. CAPAN2	4.6
Renal ca. UO-31	4.5	Pancreas Pool	7.3

Table TD. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5023, Run 223740941	Tissue Name	Rel. Exp.(%) Ag5023, Run 223740941
Secondary Th1 act	80.7	HUVEC IL-1beta	20.4
Secondary Th2 act	100.0	HUVEC IFN gamma	26.4
Secondary Tr1 act	67.4	HUVEC TNF alpha + IFN gamma	20.0
Secondary Th1 rest	18.7	HUVEC TNF alpha + IL4	25.7
Secondary Th2 rest	43.8	HUVEC IL-11	16.6
Secondary Tr1 rest	10.7	Lung Microvascular EC none	46.7
Primary Th1 act	42.0	Lung Microvascular EC TNFalpha + IL-1beta	37.6
Primary Th2 act	46.3	Microvascular Dermal EC none	33.0
Primary Tr1 act	60.7	Microvascular Dermal EC TNFalpha + IL-1beta	21.9
Primary Th1 rest	22.7	Bronchial epithelium TNFalpha + IL1beta	35.1
Primary Th2 rest	6.5	Small airway epithelium none	11.0
Primary Tr1 rest	33.2	Small airway epithelium TNFalpha + IL-1beta	20.2
CD45RA CD4 lymphocyte act	47.0	Coronary artery SMC rest	6.1
CD45RO CD4 lymphocyte act	52.1	Coronary artery SMC TNFalpha + IL-1beta	24.5
CD8 lymphocyte act	48.3	Astrocytes rest	12.5
Secondary CD8 lymphocyte rest	42.9	Astrocytes TNFalpha + IL-1beta	8.5
Secondary CD8 lymphocyte act	24.5	KU-812 (Basophil) rest	32.1
CD4 lymphocyte none	10.5	KU-812 (Basophil) PMA/ionomycin	43.2

2ry Th1/Th2/Tr1_anti-CD95 CHI1	27.9	CCD1106 (Keratinocytes) none	39.2
LAK cells rest	23.0	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	23.8
LAK cells IL-2	60.3	Liver cirrhosis	4.9
LAK cells IL-2+IL-12	25.9	NCI-H292 none	22.8
LAK cells IL-2+IFN gamma	28.5	NCI-H292 IL-4	27.2
LAK cells IL-2+ IL-18	32.3	NCI-H292 IL-9	41.2
LAK cells PMA/ionomycin	19.8	NCI-H292 IL-13	40.9
NK Cells IL-2 rest	44.8	NCI-H292 IFN gamma	45.4
Two Way MLR 3 day	27.7	HPAEC none	17.0
Two Way MLR 5 day	28.7	HPAEC TNF alpha + IL-1 beta	40.9
Two Way MLR 7 day	28.7	Lung fibroblast none	23.2
PBMC rest	15.2	Lung fibroblast TNF alpha + IL-1 beta	20.3
PBMC PWM	55.5	Lung fibroblast IL-4	34.2
PBMC PHA-L	31.4	Lung fibroblast IL-9	40.1
Ramos (B cell) none	91.4	Lung fibroblast IL-13	25.3
Ramos (B cell) ionomycin	46.7	Lung fibroblast IFN gamma	39.8
B lymphocytes PWM	34.4	Dermal fibroblast CCD1070 rest	49.0
B lymphocytes CD40L and IL-4	47.3	Dermal fibroblast CCD1070 TNF alpha	63.7
EOL-1 dbcAMP	50.7	Dermal fibroblast CCD1070 IL-1 beta	38.4
EOL-1 dbcAMP PMA/ionomycin	51.8	Dermal fibroblast IFN gamma	18.8
Dendritic cells none	32.5	Dermal fibroblast IL-4	44.4
Dendritic cells LPS	11.2	Dermal Fibroblasts rest	24.8
Dendritic cells anti-CD40	29.1	Neutrophils TNFa+LPS	0.0
Monocytes rest	11.1	Neutrophils rest	3.5
Monocytes LPS	12.2	Colon	5.0
Macrophages rest	30.8	Lung	5.9
Macrophages LPS	8.1	Thymus	20.4
HUVEC none	19.9	Kidney	24.7
HUVEC starved	31.0		

CNS_neurodegeneration_v1.0 Summary: Ag5023 This gene appears to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

General_screening_panel_v1.5 Summary: Ag5023 This gene is widely expressed in this panel, with highest expression in fetal liver (CT=26.8). In addition, this gene is expressed at much higher levels in fetal lung and lung tissue (CT=30) when compared to expression in the adult counterparts (CTs=40). Thus, expression of this gene may be used to

differentiate between the fetal and adult sources of these tissues. In addition, therapeutic modulation of the expression or function of this gene may be useful in the treatment of diseases that affect these organs.

Moderate to low expression is seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Panel 4.1D Summary: Ag5023 Highest expression of this gene is seen in chronically activated Th2 cells (CT=30). This gene is also expressed at moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.5 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

U. NOV78a: Selenoprotein X 1

Expression of gene NOV78a was assessed using the primer-probe set Ag5042, described in Table UA.

Table UA. Probe Name Ag5042

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgagccacaagttcctgaac-3'	20	263	408
Probe	TET-5'-agtcgccgattcatattcagcagctcg-3'-TAMRA	26	302	409
Reverse	5'-tgcctttagggacaaactca-3'	21	329	410

V. NOV79a: Hypothetical WD-repeat

Expression of gene NOV79a was assessed using the primer-probe set Ag5050, described in Table VA. Results of the RTQ-PCR runs are shown in Tables VB and VC.

Table VA. Probe Name Ag5050

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtgaagccaaactatgctctca-3'	22	121	411
Probe	TET-5'-caaagcagtgctccgtgaaattca-3'-TAMRA	26	165	412
Reverse	5'-atgaacttgccagccactct-3'	20	201	413

Table VB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag5050, Run 224080134	Tissue Name	Rel. Exp.(%) Ag5050, Run 224080134
AD 1 Hippo	25.3	Control (Path) 3 Temporal Ctx	5.2
AD 2 Hippo	27.2	Control (Path) 4 Temporal Ctx	34.9
AD 3 Hippo	8.9	AD 1 Occipital Ctx	14.5
AD 4 Hippo	4.6	AD 2 Occipital Ctx (Missing)	0.0
AD 5 Hippo	98.6	AD 3 Occipital Ctx	9.4
AD 6 Hippo	67.4	AD 4 Occipital Ctx	9.1
Control 2 Hippo	23.5	AD 5 Occipital Ctx	47.3
Control 4 Hippo	13.2	AD 6 Occipital Ctx	29.5
Control (Path) 3 Hippo	8.0	Control 1 Occipital Ctx	4.5
AD 1 Temporal Ctx	20.3	Control 2 Occipital Ctx	50.7
AD 2 Temporal Ctx	30.6	Control 3 Occipital Ctx	16.5
AD 3 Temporal Ctx	6.7	Control 4 Occipital Ctx	6.9

AD 4 Temporal Ctx	23.5	Control (Path) 1 Occipital Ctx	70.2
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	10.2
AD 5 Sup Temporal Ctx	51.4	Control (Path) 3 Occipital Ctx	5.4
AD 6 Inf Temporal Ctx	74.7	Control (Path) 4 Occipital Ctx	17.2
AD 6 Sup Temporal Ctx	85.9	Control 1 Parietal Ctx	6.9
Control 1 Temporal Ctx	7.5	Control 2 Parietal Ctx	42.9
Control 2 Temporal Ctx	47.6	Control 3 Parietal Ctx	21.9
Control 3 Temporal Ctx	17.9	Control (Path) 1 Parietal Ctx	68.8
Control 3 Temporal Ctx	12.2	Control (Path) 2 Parietal Ctx	22.2
Control (Path) 1 Temporal Ctx	55.9	Control (Path) 3 Parietal Ctx	6.4
Control (Path) 2 Temporal Ctx	34.2	Control (Path) 4 Parietal Ctx	36.9

Table VC. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag5050, Run 223796098	Tissue Name	Rel. Exp.(%) Ag5050, Run 223796098
Secondary Th1 act	59.0	HUVEC IL-1beta	44.4
Secondary Th2 act	63.7	HUVEC IFN gamma	33.7
Secondary Tr1 act	49.0	HUVEC TNF alpha + IFN gamma	26.4
Secondary Th1 rest	7.4	HUVEC TNF alpha + IL4	52.1
Secondary Th2 rest	12.4	HUVEC IL-11	19.1
Secondary Tr1 rest	9.5	Lung Microvascular EC none	49.3
Primary Th1 act	45.1	Lung Microvascular EC TNFalpha + IL-1beta	36.3
Primary Th2 act	63.7	Microvascular Dermal EC none	23.5
Primary Tr1 act	54.7	Microvascular Dermal EC TNFalpha + IL-1beta	25.3
Primary Th1 rest	11.0	Bronchial epithelium TNFalpha + IL1beta	28.7
Primary Th2 rest	1.9	Small airway epithelium none	15.5
Primary Tr1 rest	15.8	Small airway epithelium TNFalpha + IL-1beta	23.0
CD45RA CD4 lymphocyte act	49.3	Coronary artery SMC rest	18.4
CD45RO CD4 lymphocyte act	44.8	Coronary artery SMC TNFalpha + IL-1beta	13.0
CD8 lymphocyte act	51.4	Astrocytes rest	7.2
Secondary CD8 lymphocyte rest	44.1	Astrocytes TNFalpha + IL-1beta	7.0
Secondary CD8 lymphocyte act	15.0	KU-812 (Basophil) rest	44.8
CD4 lymphocyte none	6.3	KU-812 (Basophil) PMA/ionomycin	48.3
2ry Th1/Th2/Tr1_anti-CD95 CH11	13.6	CCD1106 (Keratinocytes) none	100.0

LAK cells rest	10.9	CCD1106 (Keratinocytes) TNFalpha + IL-1beta	45.1
LAK cells IL-2	26.4	Liver cirrhosis	4.5
LAK cells IL-2+IL-12	16.4	NCI-H292 none	24.5
LAK cells IL-2+IFN gamma	17.7	NCI-H292 IL-4	48.6
LAK cells IL-2+ IL-18	20.3	NCI-H292 IL-9	33.7
LAK cells PMA/ionomycin	24.8	NCI-H292 IL-13	40.3
NK Cells IL-2 rest	14.9	NCI-H292 IFN gamma	46.0
Two Way MLR 3 day	22.7	HPAEC none	15.1
Two Way MLR 5 day	31.0	HPAEC TNF alpha + IL-1 beta	55.5
Two Way MLR 7 day	18.7	Lung fibroblast none	27.5
PBMC rest	5.8	Lung fibroblast TNF alpha + IL-1 beta	14.1
PBMC PWM	35.4	Lung fibroblast IL-4	44.4
PBMC PHA-L	29.1	Lung fibroblast IL-9	42.6
Ramos (B cell) none	46.0	Lung fibroblast IL-13	25.9
Ramos (B cell) ionomycin	53.6	Lung fibroblast IFN gamma	45.4
B lymphocytes PWM	35.1	Dermal fibroblast CCD1070 rest	54.7
B lymphocytes CD40L and IL-4	19.5	Dermal fibroblast CCD1070 TNF alpha	50.0
EOL-1 dbcAMP	66.4	Dermal fibroblast CCD1070 IL-1 beta	28.7
EOL-1 dbcAMP PMA/ionomycin	44.4	Dermal fibroblast IFN gamma	14.9
Dendritic cells none	13.9	Dermal fibroblast IL-4	35.6
Dendritic cells LPS	10.2	Dermal Fibroblasts rest	17.0
Dendritic cells anti-CD40	21.2	Neutrophils TNFa+LPS	0.2
Monocytes rest	11.5	Neutrophils rest	1.3
Monocytes LPS	15.0	Colon	5.7
Macrophages rest	14.8	Lung	7.2
Macrophages LPS	5.3	Thymus	15.1
HUVEC none	32.5	Kidney	12.8
HUVEC starved	39.5		

CNS_neurodegeneration_v1.0 Summary: Ag5050 This gene appears to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

Panel 4.1D Summary: Ag5050 Highest expression of this gene is seen in keratinocytes (CT=29). This gene is also expressed at moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin,

and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

W. NOV84a and NOV84b: GTF2IRD1

Expression of gene NOV84a and full-length physical clone NOV84b was assessed using the primer-probe set Ag3588, described in Table WA. Results of the RTQ-PCR runs are shown in Tables WB and WC. Please note that NOV84b represents a full-length physical clone of the NOV84a gene, validating the prediction of the gene sequence.

Table WA. Probe Name Ag3588

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tgggagagagcgtattcttc-3'	21	1670	414
Probe	TET-5'-tggaagtacagaatattccaacatgtctca-3'-TAMRA	30	1639	415
Reverse	5'-acacagacatgcttgtttgc-3'	21	1615	416

Table WB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag3588, Run 211006685	Tissue Name	Rel. Exp.(%) Ag3588, Run 211006685
AD 1 Hippo	24.8	Control (Path) 3 Temporal Ctx	10.0
AD 2 Hippo	35.8	Control (Path) 4 Temporal Ctx	25.0
AD 3 Hippo	12.5	AD 1 Occipital Ctx	21.8
AD 4 Hippo	13.5	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	55.1	AD 3 Occipital Ctx	12.8
AD 6 Hippo	50.3	AD 4 Occipital Ctx	26.2
Control 2 Hippo	29.9	AD 5 Occipital Ctx	12.8
Control 4 Hippo	22.2	AD 6 Occipital Ctx	33.7
Control (Path) 3 Hippo	9.7	Control 1 Occipital Ctx	5.0
AD 1 Temporal Ctx	35.4	Control 2 Occipital Ctx	45.4
AD 2 Temporal Ctx	32.3	Control 3 Occipital Ctx	24.5
AD 3 Temporal Ctx	8.7	Control 4 Occipital Ctx	15.1

AD 4 Temporal Ctx	28.7	Control (Path) 1 Occipital Ctx	77.4
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	13.5
AD 5 Sup Temporal Ctx	50.3	Control (Path) 3 Occipital Ctx	3.6
AD 6 Inf Temporal Ctx	58.2	Control (Path) 4 Occipital Ctx	15.3
AD 6 Sup Temporal Ctx	44.1	Control 1 Parietal Ctx	12.9
Control 1 Temporal Ctx	9.6	Control 2 Parietal Ctx	49.3
Control 2 Temporal Ctx	28.7	Control 3 Parietal Ctx	14.9
Control 3 Temporal Ctx	20.9	Control (Path) 1 Parietal Ctx	52.5
Control 4 Temporal Ctx	16.4	Control (Path) 2 Parietal Ctx	25.2
Control (Path) 1 Temporal Ctx	46.7	Control (Path) 3 Parietal Ctx	7.5
Control (Path) 2 Temporal Ctx	28.3	Control (Path) 4 Parietal Ctx	43.8

Table WC. general oncology screening panel v 2.4

Tissue Name	Rel. Exp.(%) Ag3588, Run 267747154	Tissue Name	Rel. Exp.(%) Ag3588, Run 267747154
Colon cancer 1	8.2	Bladder NAT 2	1.3
Colon NAT 1	5.8	Bladder NAT 3	1.0
Colon cancer 2	4.4	Bladder NAT 4	10.4
Colon NAT 2	8.2	Prostate adenocarcinoma 1	32.8
Colon cancer 3	18.3	Prostate adenocarcinoma 2	6.5
Colon NAT 3	17.3	Prostate adenocarcinoma 3	15.1
Colon malignant cancer 4	8.5	Prostate adenocarcinoma 4	13.0
Colon NAT 4	2.6	Prostate NAT 5	4.7
Lung cancer 1	13.0	Prostate adenocarcinoma 6	7.3
Lung NAT 1	2.4	Prostate adenocarcinoma 7	8.5
Lung cancer 2	37.4	Prostate adenocarcinoma 8	3.3
Lung NAT 2	2.2	Prostate adenocarcinoma 9	36.9
Squamous cell carcinoma 3	12.5	Prostate NAT 10	3.9
Lung NAT 3	1.5	Kidney cancer 1	25.5
Metastatic melanoma 1	56.3	Kidney NAT 1	15.8
Melanoma 2	2.5	Kidney cancer 2	35.8
Melanoma 3	2.8	Kidney NAT 2	25.9
Metastatic melanoma 4	88.3	Kidney cancer 3	36.9
Metastatic melanoma 5	100.0	Kidney NAT 3	10.6
Bladder cancer 1	2.5	Kidney cancer 4	13.9
Bladder NAT 1	0.0	Kidney NAT 4	6.1
Bladder cancer 2	4.2		

5 **CNS_neurodegeneration_v1.0 Summary:** Ag3588 This gene appears to be slightly upregulated in the temporal cortex of Alzheimer's disease patients. Therefore, therapeutic

modulation of the expression or function of this gene may decrease neuronal death and be of use in the treatment of this disease.

general oncology screening panel_v_2.4 Summary: Ag3588 This gene is widely expressed in this panel, with highest expression in melanoma (CT=28.6). In addition, this gene is more highly expressed in lung cancer than in the corresponding normal adjacent tissue. Thus, expression of this gene could be used as a marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of lung and melanoma cancer.

X. NOV85a and NOV85b: Intracellular Protein

Expression of gene NOV85a and full-length physical clone NOV85b was assessed using the primer-probe sets Ag3597 and Ag3679, described in Tables XA and XB. Results of the RTQ-PCR runs are shown in Tables XC, XD, XE and XF. Please note that NOV85b represents a full-length physical clone of the NOV85a gene, validating the prediction of the gene sequence.

Table XA. Probe Name Ag3597

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-aaggaacacagcctactgtca-3'	22	187	417
Probe	TET-5'-cttcaaccacctaacagccacagcag-3'-TAMRA	26	158	418
Reverse	5'-aaagcccactaggagagagaca-3'	22	132	419

Table XB. Probe Name Ag3679

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-acaaaggaacacagcctacttg-3'	22	190	420
Probe	TET-5'-cttcaaccacctaacagccacagcag-3'-TAMRA	26	158	421
Reverse	5'-gccactaggagagagacactt-3'	22	135	422

Table XC. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag3597, Run 211010103	Tissue Name	Rel. Exp.(%) Ag3597, Run 211010103
AD 1 Hippo	18.2	Control (Path) 3 Temporal Ctx	11.0
AD 2 Hippo	24.0	Control (Path) 4 Temporal Ctx	28.7
AD 3 Hippo	13.8	AD 1 Occipital Ctx	21.0

AD 4 Hippo	7.1	AD 2 Occipital Ctx (Missing)	0.0
AD 5 hippo	72.7	AD 3 Occipital Ctx	11.3
AD 6 Hippo	47.6	AD 4 Occipital Ctx	18.0
Control 2 Hippo	19.5	AD 5 Occipital Ctx	32.3
Control 4 Hippo	9.8	AD 6 Occipital Ctx	30.8
Control (Path) 3 Hippo	11.3	Control 1 Occipital Ctx	7.9
AD 1 Temporal Ctx	26.6	Control 2 Occipital Ctx	33.0
AD 2 Temporal Ctx	32.3	Control 3 Occipital Ctx	18.7
AD 3 Temporal Ctx	7.0	Control 4 Occipital Ctx	8.8
AD 4 Temporal Ctx	29.1	Control (Path) 1 Occipital Ctx	55.9
AD 5 Inf Temporal Ctx	100.0	Control (Path) 2 Occipital Ctx	12.5
AD 5 Sup Temporal Ctx	49.7	Control (Path) 3 Occipital Ctx	11.3
AD 6 Inf Temporal Ctx	47.0	Control (Path) 4 Occipital Ctx	14.6
AD 6 Sup Temporal Ctx	42.9	Control 1 Parietal Ctx	13.1
Control 1 Temporal Ctx	10.0	Control 2 Parietal Ctx	54.0
Control 2 Temporal Ctx	25.2	Control 3 Parietal Ctx	15.9
Control 3 Temporal Ctx	17.1	Control (Path) 1 Parietal Ctx	43.2
Control 4 Temporal Ctx	12.7	Control (Path) 2 Parietal Ctx	22.4
Control (Path) 1 Temporal Ctx	37.9	Control (Path) 3 Parietal Ctx	10.7
Control (Path) 2 Temporal Ctx	27.9	Control (Path) 4 Parietal Ctx	28.3

Table XD. General screening panel v1.4

Tissue Name	Rel. Exp.(%) Ag3597, Run 218307127	Rel. Exp.(%) Ag3679, Run 218941309	Tissue Name	Rel. Exp.(%) Ag3597, Run 218307127	Rel. Exp.(%) Ag3679, Run 218941309
Adipose	17.7	4.6	Renal ca. TK-10	26.8	25.0
Melanoma* Hs688(A).T	22.2	22.5	Bladder	23.0	27.7
Melanoma* Hs688(B).T	22.1	23.5	Gastric ca. (liver met.) NCI-N87	36.9	37.1
Melanoma* M14	19.9	21.3	Gastric ca. KATO III	45.7	51.8
Melanoma* LOXIMVI	23.3	21.5	Colon ca. SW-948	7.2	10.7
Melanoma* SK- MEL-5	27.4	38.2	Colon ca. SW480	26.8	46.0
Squamous cell carcinoma SCC-4	22.8	32.3	Colon ca.* (SW480 met) SW620	21.5	19.3
Testis Pool	31.0	26.1	Colon ca. HT29	11.4	10.5
Prostate ca.* (bone met) PC-3	42.3	43.5	Colon ca. HCT-116	32.1	34.9
Prostate Pool	12.4	13.1	Colon ca. CaCo-2	27.7	33.9
Placenta	20.3	21.0	Colon cancer tissue	15.6	12.8

Uterus Pool	13.1	12.8	Colon ca. SW1116	11.0	12.5
Ovarian ca. OVCAR-3	33.2	26.8	Colon ca. Colo-205	2.9	5.1
Ovarian ca. SK-OV-3	36.6	25.5	Colon ca. SW-48	3.8	6.5
Ovarian ca. OVCAR-4	16.4	17.8	Colon Pool	19.2	24.3
Ovarian ca. OVCAR-5	33.2	58.2	Small Intestine Pool	31.4	33.7
Ovarian ca. IGROV-1	15.7	18.6	Stomach Pool	11.9	14.8
Ovarian ca. OVCAR-8	5.8	9.7	Bone Marrow Pool	11.2	10.6
Ovary	13.4	12.1	Fetal Heart	22.7	24.3
Breast ca. MCF-7	47.0	57.8	Heart Pool	10.5	12.8
Breast ca. MDA-MB-231	40.9	48.3	Lymph Node Pool	27.0	24.7
Breast ca. BT 549	52.1	50.0	Fetal Skeletal Muscle	17.3	18.3
Breast ca. T47D	100.0	100.0	Skeletal Muscle Pool	30.8	28.5
Breast ca. MDA-N	13.6	21.8	Spleen Pool	17.1	19.9
Breast Pool	22.2	20.7	Thymus Pool	19.8	19.3
Trachea	21.5	21.2	CNS cancer (glio/astro) U87-MG	31.9	45.4
Lung	5.6	5.3	CNS cancer (glio/astro) U-118-MG	46.3	56.3
Fetal Lung	36.9	35.6	CNS cancer (neuro;met) SK-N-AS	29.3	27.4
Lung ca. NCI-N417	4.0	7.3	CNS cancer (astro) SF-539	10.4	12.5
Lung ca. LX-1	36.1	34.6	CNS cancer (astro) SNB-75	40.1	51.4
Lung ca. NCI-H146	5.3	6.3	CNS cancer (glio) SNB-19	14.2	19.9
Lung ca. SHP-77	13.5	24.5	CNS cancer (glio) SF-295	49.7	44.4
Lung ca. A549	22.5	27.7	Brain (Amygdala) Pool	22.5	20.7
Lung ca. NCI-H526	8.4	12.5	Brain (cerebellum)	77.9	79.0
Lung ca. NCI-H23	24.0	35.1	Brain (fetal)	36.9	37.1
Lung ca. NCI-H460	9.9	15.5	Brain (Hippocampus) Pool	18.8	19.9
Lung ca. HOP-62	9.6	12.8	Cerebral Cortex Pool	19.9	21.3
Lung ca. NCI-H522	24.0	23.8	Brain (Substantia nigra) Pool	18.9	19.6
Liver	5.1	5.2	Brain (Thalamus) Pool	30.1	31.0
Fetal Liver	14.0	24.1	Brain (whole)	23.8	25.5
Liver ca. HepG2	17.0	18.8	Spinal Cord Pool	25.0	27.7
Kidney Pool	30.8	43.2	Adrenal Gland	36.1	34.6
Fetal Kidney	24.7	28.5	Pituitary gland Pool	5.8	8.2
Renal ca. 786-0	21.8	21.6	Salivary Gland	15.6	15.7

Renal ca. A498	4.2	4.6	Thyroid (female)	13.4	13.2
Renal ca. ACHN	17.6	18.8	Pancreatic ca. CAPAN2	27.7	27.5
Renal ca. UO-31	24.7	22.1	Pancreas Pool	29.7	27.9

Table XE. Panel 4.1D

Tissue Name	Rel. Exp.(%) Ag3597, Run 169910426	Rel. Exp.(%) Ag3679, Run 169988037	Tissue Name	Rel. Exp.(%) Ag3597, Run 169910426	Rel. Exp.(%) Ag3679, Run 169988037
Secondary Th1 act	63.7	64.6	HUVEC IL-1beta	25.9	18.8
Secondary Th2 act	64.2	95.3	HUVEC IFN gamma	34.4	33.0
Secondary Tr1 act	82.4	87.7	HUVEC TNF alpha + IFN gamma	25.3	27.5
Secondary Th1 rest	26.8	41.8	HUVEC TNF alpha + IL4	27.2	30.4
Secondary Th2 rest	42.3	60.7	HUVEC IL-11	13.1	21.6
Secondary Tr1 rest	36.6	46.0	Lung Microvascular EC none	44.4	52.1
Primary Th1 act	43.5	54.0	Lung Microvascular EC TNFalpha + IL-1beta	48.3	48.6
Primary Th2 act	55.5	63.3	Microvascular Dermal EC none	24.3	35.1
Primary Tr1 act	51.1	73.7	Microvascular Dermal EC TNFalpha + IL- 1beta	25.9	24.8
Primary Th1 rest	48.6	56.3	Bronchial epithelium TNFalpha + IL1beta	35.4	31.9
Primary Th2 rest	46.7	57.4	Small airway epithelium none	17.2	18.7
Primary Tr1 rest	49.7	69.3	Small airway epithelium TNFalpha + IL-1beta	38.2	46.3
CD45RA CD4 lymphocyte act	51.4	63.3	Coronary artery SMC rest	24.0	36.6
CD45RO CD4 lymphocyte act	66.9	95.3	Coronary artery SMC TNFalpha + IL-1beta	33.0	32.5
CD8 lymphocyte act	58.6	75.8	Astrocytes rest	19.8	26.6
Secondary CD8 lymphocyte rest	51.1	69.3	Astrocytes TNFalpha + IL-1beta	17.2	26.6
Secondary CD8 lymphocyte act	38.7	37.9	KU-812 (Basophil) rest	37.1	50.7
CD4 lymphocyte none	42.0	58.6	KU-812 (Basophil) PMA/ionomycin	72.7	68.3
2ry Th1/Th2/Tr1_anti- CD95 CH11	41.5	56.6	CCD1106 (Keratinocytes) none	65.1	64.2

LAK cells rest	61.1	71.7	CCD1106 (Keratinocytes) TNFalpha + IL-1 beta	48.3	58.6
LAK cells IL-2	61.1	72.7	Liver cirrhosis	14.9	20.4
LAK cells IL-2+IL-12	100.0	62.0	NCI-H292 none	22.1	30.6
LAK cells IL-2+IFN gamma	85.3	65.1	NCI-H292 IL-4	36.9	42.0
LAK cells IL-2+ IL-18	73.7	100.0	NCI-H292 IL-9	62.9	70.2
LAK cells PMA/ionomycin	58.6	83.5	NCI-H292 IL-13	42.0	37.4
NK Cells IL-2 rest	59.9	98.6	NCI-H292 IFN gamma	46.0	48.3
Two Way MLR 3 day	72.7	65.5	HPAEC none	27.4	26.2
Two Way MLR 5 day	43.8	56.3	HPAEC TNF alpha + IL-1 beta	37.1	48.3
Two Way MLR 7 day	29.3	40.1	Lung fibroblast none	27.0	29.5
PBMC rest	44.1	58.6	Lung fibroblast TNF alpha + IL-1 beta	17.2	24.7
PBMC PWM	48.3	60.7	Lung fibroblast IL-4	25.3	31.6
PBMC PHA-L	31.9	52.5	Lung fibroblast IL-9	45.4	43.2
Ramos (B cell) none	65.5	87.1	Lung fibroblast IL-13	30.1	25.0
Ramos (B cell) ionomycin	71.2	87.1	Lung fibroblast IFN gamma	31.4	32.1
B lymphocytes PWM	33.2	52.9	Dermal fibroblast CCD1070 rest	45.4	51.1
B lymphocytes CD40L and IL-4	58.2	78.5	Dermal fibroblast CCD1070 TNF alpha	74.2	98.6
EOL-1 dbcAMP	40.1	60.3	Dermal fibroblast CCD1070 IL-1 beta	32.5	34.9
EOL-1 dbcAMP PMA/ionomycin	50.7	75.8	Dermal fibroblast IFN gamma	20.3	27.9
Dendritic cells none	41.5	52.9	Dermal fibroblast IL-4	41.2	41.2
Dendritic cells LPS	28.1	42.0	Dermal Fibroblasts rest	24.8	29.7
Dendritic cells anti- CD40	36.9	40.9	Neutrophils TNFa+LPS	15.6	29.5
Monocytes rest	55.1	60.3	Neutrophils rest	84.1	76.8
Monocytes LPS	57.4	82.4	Colon	34.9	34.4
Macrophages rest	40.1	54.0	Lung	31.0	29.3
Macrophages LPS	22.5	31.4	Thymus	90.1	85.3
HUVEC none	15.0	24.0	Kidney	49.7	52.5
HUVEC starved	28.1	29.7			

Table XF. general oncology screening panel v 2.4

Tissue Name	Rel. Exp.(%) Ag3597, Run 267747376	Rel. Exp.(%) Ag3679, Run 267742157	Tissue Name	Rel. Exp.(%) Ag3597, Run 267747376	Rel. Exp.(%) Ag3679, Run 267742157
Colon cancer 1	20.4	28.1	Bladder NAT 2	1.6	2.1
Colon NAT 1	24.5	13.2	Bladder NAT 3	1.7	2.2
Colon cancer 2	32.5	28.7	Bladder NAT 4	17.4	13.7
Colon NAT 2	20.0	22.8	Prostate adenocarcinoma 1	50.3	69.7
Colon cancer 3	45.4	40.6	Prostate adenocarcinoma 2	8.1	8.0
Colon NAT 3	42.9	34.4	Prostate adenocarcinoma 3	27.2	20.2
Colon malignant cancer 4	75.3	73.7	Prostate adenocarcinoma 4	32.5	33.2
Colon NAT 4	11.6	11.1	Prostate NAT 5	7.0	7.3
Lung cancer 1	19.8	18.6	Prostate adenocarcinoma 6	11.9	9.6
Lung NAT 1	3.8	3.7	Prostate adenocarcinoma 7	9.3	10.5
Lung cancer 2	59.5	66.9	Prostate adenocarcinoma 8	4.7	4.5
Lung NAT 2	3.7	4.6	Prostate adenocarcinoma 9	38.7	28.9
Squamous cell carcinoma 3	55.5	47.3	Prostate NAT 10	4.8	4.9
Lung NAT 3	2.0	1.8	Kidney cancer 1	31.6	37.1
Metastatic melanoma 1	36.3	38.4	Kidney NAT 1	14.7	17.6
Melanoma 2	9.9	10.7	Kidney cancer 2	100.0	82.9
Melanoma 3	12.2	11.0	Kidney NAT 2	41.8	31.9
Metastatic melanoma 4	89.5	92.7	Kidney cancer 3	45.4	34.9
Metastatic melanoma 5	95.3	100.0	Kidney NAT 3	11.4	12.1
Bladder cancer 1	4.7	3.1	Kidney cancer 4	41.2	30.1
Bladder NAT 1	0.0	0.0	Kidney NAT 4	21.8	12.8
Bladder cancer 2	12.0	10.5			

General_screening_panel_v1.4 Summary: Ag3597/Ag3679 Two experiments with the same probe and primer produce results that are in excellent agreement. Highest expression of this gene is seen in a breast cancer cell line. Higher levels of expression are also

5 seen in breast, prostate, ovarian and lung tissues when compared to expression in normal

tissue. Thus, expression of this gene could be used as a marker of these cancers and therapeutic modulation of the activity of this gene may be effective in their treatment.

Among tissues with metabolic function, this gene is expressed at high to moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at high to moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

This gene codes for variant of DMR protein and a homologue of mouse dystrophin myotonia-containing WD repeat motif protein (DMR-N9 protein). DMR-N9 has been implicated in myotonic dystrophy (MD). Therefore, therapeutic modulation of this gene could be useful in the treatment of MD

Panel 4.1D Summary: Ag3597/Ag3679 Two experiments with the same probe and primer produce results that are in excellent agreement. Highest expression of this gene is seen in cytokine activated LAK cells. In addition, this gene is expressed at high to moderate levels in a wide range of cell types of significance in the immune response in health and disease. These cells include members of the T-cell, B-cell, endothelial cell, macrophage/monocyte, and peripheral blood mononuclear cell family, as well as epithelial and fibroblast cell types from lung and skin, and normal tissues represented by colon, lung, thymus and kidney. This ubiquitous pattern of expression suggests that this gene product may be involved in homeostatic processes for these and other cell types and tissues. This pattern is in agreement with the expression profile in General_screening_panel_v1.4 and also suggests a role for the gene product in cell survival and proliferation. Therefore, modulation of the gene product with a functional therapeutic may lead to the alteration of functions associated with these cell types and lead to improvement of the symptoms of patients suffering from autoimmune and inflammatory diseases such as asthma, allergies, inflammatory bowel disease, lupus erythematosus, psoriasis, rheumatoid arthritis, and osteoarthritis.

general oncology screening panel_v_2.4 Summary: Ag3597/Ag3679 Two experiments produce results that are in very good agreement, with highest expression in a kidney cancer and a melanoma (CT=27.3-28.7). In addition, this gene is more highly expressed in lung and kidney cancer than in the corresponding normal adjacent tissue. In addition, consistent prominent expression is seen in melanoma and prostate cancer. Thus, expression of this gene could be used as a marker of these cancers. Furthermore, therapeutic modulation of the expression or function of this gene product may be useful in the treatment of lung, melanoma, prostate and kidney cancer.

10 Y. NOV87a and NOV87b: Glycolipid Transfer Protein-like

Expression of gene NOV87a and full-length physical clone NOV87b was assessed using the primer-probe set Ag6896, described in Table YA. Results of the RTQ-PCR runs are shown in Table YB. Please note that NOV87b represents a full-length physical clone of the NOV87a gene, validating the prediction of the gene sequence.

15 Table YA. Probe Name Ag6896

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-cgtcacccgtggccttct-3'	17	741	423
Probe	TET-5'-cacgctgcccacacgcgagg-3'-TAMRA	20	759	424
Reverse	5'-gttcattggcctccaggaaga-3'	20	779	425

Table YB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6896, Run 278388389	Tissue Name	Rel. Exp.(%) Ag6896, Run 278388389
Adipose	6.8	Renal ca. TK-10	19.6
Melanoma* Hs688(A).T	33.9	Bladder	17.0
Melanoma* Hs688(B).T	29.1	Gastric ca. (liver met.) NCI-N87	69.3
Melanoma* M14	50.7	Gastric ca. KATO III	55.1
Melanoma* LOXIMVI	20.9	Colon ca. SW-948	64.2
Melanoma* SK-MEL-5	20.6	Colon ca. SW480	65.5
Squamous cell carcinoma SCC-4	53.2	Colon ca.* (SW480 met) SW620	17.2
Testis Pool	35.1	Colon ca. HT29	22.1
Prostate ca.* (bone met) PC-3	97.9	Colon ca. HCT-116	36.6
Prostate Pool	9.5	Colon ca. CaCo-2	26.1
Placenta	26.2	Colon cancer tissue	33.0

Uterus Pool	4.1	Colon ca. SW1116	17.9
Ovarian ca. OVCAR-3	99.3	Colon ca. Colo-205	17.4
Ovarian ca. SK-OV-3	94.0	Colon ca. SW-48	25.0
Ovarian ca. OVCAR-4	58.6	Colon Pool	19.1
Ovarian ca. OVCAR-5	60.3	Small Intestine Pool	15.6
Ovarian ca. IGROV-1	57.8	Stomach Pool	8.7
Ovarian ca. OVCAR-8	66.9	Bone Marrow Pool	6.9
Ovary	8.8	Fetal Heart	18.8
Breast ca. MCF-7	36.3	Heart Pool	16.2
Breast ca. MDA-MB-231	100.0	Lymph Node Pool	14.6
Breast ca. BT 549	79.0	Fetal Skeletal Muscle	10.3
Breast ca. T47D	18.8	Skeletal Muscle Pool	15.9
Breast ca. MDA-N	60.3	Spleen Pool	16.2
Breast Pool	17.8	Thymus Pool	19.3
Trachea	28.7	CNS cancer (glio/astro) U87-MG	47.6
Lung	1.4	CNS cancer (glio/astro) U-118-MG	75.8
Fetal Lung	21.9	CNS cancer (neuro;met) SK-N-AS	38.4
Lung ca. NCI-N417	33.7	CNS cancer (astro) SF-539	34.6
Lung ca. LX-1	22.2	CNS cancer (astro) SNB-75	79.6
Lung ca. NCI-H146	13.3	CNS cancer (glio) SNB-19	60.7
Lung ca. SHP-77	33.2	CNS cancer (glio) SF-295	40.6
Lung ca. A549	24.5	Brain (Amygdala) Pool	20.0
Lung ca. NCI-H526	27.0	Brain (cerebellum)	12.5
Lung ca. NCI-H23	32.5	Brain (fetal)	40.1
Lung ca. NCI-H460	9.9	Brain (Hippocampus) Pool	18.2
Lung ca. HOP-62	15.8	Cerebral Cortex Pool	19.1
Lung ca. NCI-H522	54.7	Brain (Substantia nigra) Pool	21.0
Liver	21.8	Brain (Thalamus) Pool	19.2
Fetal Liver	14.3	Brain (whole)	35.6
Liver ca. HepG2	19.6	Spinal Cord Pool	16.5
Kidney Pool	21.9	Adrenal Gland	35.1
Fetal Kidney	14.4	Pituitary gland Pool	3.3
Renal ca. 786-0	28.5	Salivary Gland	46.7
Renal ca. A498	17.9	Thyroid (female)	11.4
Renal ca. ACHN	18.9	Pancreatic ca. CAPAN2	51.4
Renal ca. UO-31	24.8	Pancreas Pool	11.0

General_screening_panel_v1.6 Summary: Ag6896 Highest expression of this gene is seen in a breast cancer cell line (CT=27.8). This gene is ubiquitously expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in

cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

In addition, this gene is expressed at much higher levels in fetal lung tissue (CT=29.8) when compared to expression in the adult counterpart (CT=33.8). Thus, expression of this gene may be used to differentiate between the fetal and adult source of these tissue

Among tissues with metabolic function, this gene is expressed at moderate levels in pituitary, adipose, adrenal gland, pancreas, thyroid, and adult and fetal skeletal muscle, heart, and liver. This widespread expression among these tissues suggests that this gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at moderate levels in the CNS, including the hippocampus, thalamus, substantia nigra, amygdala, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

Z. NOV88a: Copine VII

Expression of gene NOV88a was assessed using the primer-probe set Ag3641, described in Table ZA. Results of the RTQ-PCR runs are shown in Tables ZB and ZC.

Table ZA. Probe Name Ag3641

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-tggactattacaatggcaaagg-3'	22	1748	426
Probe	TET-5'-atgaatctccagcacactagcacca-3'-TAMRA	26	1799	427
Reverse	5'-gtaaaactgtgtggggagttca-3'	22	1825	428

Table ZB. CNS neurodegeneration v1.0

Tissue Name	Rel. Exp.(%) Ag3641, Run 212315202	Tissue Name	Rel. Exp.(%) Ag3641, Run 212315202
AD 1 Hippo	2.7	Control (Path) 3 Temporal Ctx	1.1
AD 2 Hippo	6.0	Control (Path) 4 Temporal Ctx	11.7
AD 3 Hippo	1.0	AD 1 Occipital Ctx	2.7
AD 4 Hippo	0.0	AD 2 Occipital Ctx (Missing)	0.0

AD 5 hippo	100.0	AD 3 Occipital Ctx	1.0
AD 6 Hippo	24.8	AD 4 Occipital Ctx	7.5
Control 2 Hippo	6.5	AD 5 Occipital Ctx	6.0
Control 4 Hippo	1.0	AD 6 Occipital Ctx	18.9
Control (Path) 3 Hippo	0.3	Control 1 Occipital Ctx	0.0
AD 1 Temporal Ctx	1.9	Control 2 Occipital Ctx	25.2
AD 2 Temporal Ctx	9.0	Control 3 Occipital Ctx	4.8
AD 3 Temporal Ctx	1.2	Control 4 Occipital Ctx	1.2
AD 4 Temporal Ctx	2.7	Control (Path) 1 Occipital Ctx	26.4
AD 5 Inf Temporal Ctx	31.6	Control (Path) 2 Occipital Ctx	1.7
AD 5 Sup Temporal Ctx	19.6	Control (Path) 3 Occipital Ctx	0.4
AD 6 Inf Temporal Ctx	7.6	Control (Path) 4 Occipital Ctx	6.0
AD 6 Sup Temporal Ctx	13.6	Control 1 Parietal Ctx	0.5
Control 1 Temporal Ctx	0.4	Control 2 Parietal Ctx	11.6
Control 2 Temporal Ctx	6.1	Control 3 Parietal Ctx	4.5
Control 3 Temporal Ctx	4.1	Control (Path) 1 Parietal Ctx	15.7
Control 4 Temporal Ctx	1.3	Control (Path) 2 Parietal Ctx	12.7
Control (Path) 1 Temporal Ctx	26.6	Control (Path) 3 Parietal Ctx	0.2
Control (Path) 2 Temporal Ctx	11.7	Control (Path) 4 Parietal Ctx	15.7

Table ZC. General screening panel v1.4

Tissue Name	Rel. Exp.(%) Ag3641, Run 218306189	Tissue Name	Rel. Exp.(%) Ag3641, Run 218306189
Adipose	0.0	Renal ca. TK-10	0.0
Melanoma* Hs688(A).T	1.1	Bladder	0.0
Melanoma* Hs688(B).T	1.2	Gastric ca. (liver met.) NCI-N87	1.7
Melanoma* M14	0.0	Gastric ca. KATO III	0.0
Melanoma* LOXIMVI	0.0	Colon ca. SW-948	0.0
Melanoma* SK-MEL-5	0.0	Colon ca. SW480	0.0
Squamous cell carcinoma SCC-4	0.0	Colon ca.* (SW480 met) SW620	0.0
Testis Pool	4.4	Colon ca. HT29	0.0
Prostate ca.* (bone met) PC-3	1.2	Colon ca. HCT-116	0.0
Prostate Pool	100.0	Colon ca. CaCo-2	0.0
Placenta	0.0	Colon cancer tissue	0.0
Uterus Pool	0.0	Colon ca. SW1116	0.0
Ovarian ca. OVCAR-3	66.9	Colon ca. Colo-205	0.0
Ovarian ca. SK-OV-3	0.0	Colon ca. SW-48	0.0
Ovarian ca. OVCAR-4	3.0	Colon Pool	0.0
Ovarian ca. OVCAR-5	0.0	Small Intestine Pool	0.0
Ovarian ca. IGROV-1	14.0	Stomach Pool	0.0

Ovarian ca. OVCAR-8	10.4	Bone Marrow Pool	0.0
Ovary	4.8	Fetal Heart	21.5
Breast ca. MCF-7	3.8	Heart Pool	32.8
Breast ca. MDA-MB-231	1.2	Lymph Node Pool	1.5
Breast ca. BT 549	0.0	Fetal Skeletal Muscle	0.0
Breast ca. T47D	1.4	Skeletal Muscle Pool	0.0
Breast ca. MDA-N	0.0	Spleen Pool	6.4
Breast Pool	1.4	Thymus Pool	0.0
Trachea	10.7	CNS cancer (glio/astro) U87-MG	0.0
Lung	0.0	CNS cancer (glio/astro) U-118-MG	0.0
Fetal Lung	1.8	CNS cancer (neuro;met) SK-N-AS	0.0
Lung ca. NCI-N417	0.0	CNS cancer (astro) SF-539	0.0
Lung ca. LX-1	0.0	CNS cancer (astro) SNB-75	0.0
Lung ca. NCI-H146	14.0	CNS cancer (glio) SNB-19	11.3
Lung ca. SHP-77	0.0	CNS cancer (glio) SF-295	0.0
Lung ca. A549	75.3	Brain (Amygdala) Pool	54.3
Lung ca. NCI-H526	0.0	Brain (cerebellum)	0.0
Lung ca. NCI-H23	4.9	Brain (fetal)	54.0
Lung ca. NCI-H460	5.1	Brain (Hippocampus) Pool	59.5
Lung ca. HOP-62	0.0	Cerebral Cortex Pool	82.4
Lung ca. NCI-H522	0.0	Brain (Substantia nigra) Pool	62.4
Liver	0.0	Brain (Thalamus) Pool	92.7
Fetal Liver	0.0	Brain (whole)	39.0
Liver ca. HepG2	0.0	Spinal Cord Pool	17.9
Kidney Pool	1.4	Adrenal Gland	2.6
Fetal Kidney	1.9	Pituitary gland Pool	24.5
Renal ca. 786-0	0.0	Salivary Gland	1.6
Renal ca. A498	0.0	Thyroid (female)	0.0
Renal ca. ACHN	0.0	Pancreatic ca. CAPAN2	2.7
Renal ca. UO-31	0.0	Pancreas Pool	0.0

CNS_neurodegeneration_v1.0 Summary: Ag3641 This panel does not show differential expression of this gene in Alzheimer's disease. However, this profile confirms the expression of this gene at moderate levels in the brain. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurological disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

General_screening_panel_v1.4 Summary: Ag3641 Highest expression is seen in the prostate (CT=33). Prominent expression of this gene is seen in cells lines from ovarian and lung cancer as well as all regions of the brain in this panel. Therefore, therapeutic

modulation of this gene may be useful in the treatment of prostate related diseases, as well as, lung and ovarian cancers.

In addition, moderate to low levels of expression of this gene is also seen in all the regions of brain. Therefore, therapeutic modulation of this gene product may be useful in the treatment of neurological disorders such as Alzheimer's disease, Parkinson's disease, epilepsy, multiple sclerosis, schizophrenia and depression.

AA. NOV89b: intracellular protein

Expression of full-length physical clone NOV89b was assessed using the primer-probe set Ag6911, described in Table AAA. Results of the RTQ-PCR runs are shown in Table AAB.

Table AAA. Probe Name Ag6911

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-gtattcccaaagaaggcatctt-3'	23	412	429
Probe	TET-5'-atgaaagcaaagccagtgtcccttc-3'-TAMRA	26	442	430
Reverse	5'-gcatgtacagtggccagga-3'	19	468	431

Table AAB. General screening panel v1.6

Tissue Name	Rel. Exp.(%) Ag6911, Run 278388417	Tissue Name	Rel. Exp.(%) Ag6911, Run 278388417
Adipose	9.3	Renal ca. TK-10	63.3
Melanoma* Hs688(A).T	17.1	Bladder	20.4
Melanoma* Hs688(B).T	18.7	Gastric ca. (liver met.) NCI-N87	79.6
Melanoma* M14	49.7	Gastric ca. KATO III	63.7
Melanoma* LOXIMVI	22.1	Colon ca. SW-948	22.4
Melanoma* SK-MEL-5	42.3	Colon ca. SW480	40.3
Squamous cell carcinoma SCC-4	19.2	Colon ca.* (SW480 met) SW620	24.7
Testis Pool	16.0	Colon ca. HT29	21.9
Prostate ca.* (bone met) PC-3	13.0	Colon ca. HCT-116	46.7
Prostate Pool	17.2	Colon ca. CaCo-2	33.2
Placenta	7.0	Colon cancer tissue	14.9
Uterus Pool	2.4	Colon ca. SW1116	12.3
Ovarian ca. OVCAR-3	37.9	Colon ca. Colo-205	19.5
Ovarian ca. SK-OV-3	55.5	Colon ca. SW-48	14.2
Ovarian ca. OVCAR-4	42.0	Colon Pool	15.1
Ovarian ca. OVCAR-5	41.2	Small Intestine Pool	9.3

Ovarian ca. IGROV-1	26.6	Stomach Pool	4.1
Ovarian ca. OVCAR-8	21.8	Bone Marrow Pool	4.2
Ovary	5.6	Fetal Heart	6.4
Breast ca. MCF-7	17.3	Heart Pool	8.7
Breast ca. MDA-MB-231	66.9	Lymph Node Pool	10.6
Breast ca. BT 549	100.0	Fetal Skeletal Muscle	3.2
Breast ca. T47D	17.7	Skeletal Muscle Pool	2.7
Breast ca. MDA-N	14.7	Spleen Pool	6.5
Breast Pool	9.9	Thymus Pool	7.5
Trachea	12.6	CNS cancer (glio/astro) U87-MG	43.8
Lung	4.8	CNS cancer (glio/astro) U-118-MG	29.1
Fetal Lung	17.1	CNS cancer (neuro;met) SK-N-AS	29.3
Lung ca. NCI-N417	8.1	CNS cancer (astro) SF-539	21.0
Lung ca. LX-1	39.5	CNS cancer (astro) SNB-75	58.2
Lung ca. NCI-H146	10.2	CNS cancer (glio) SNB-19	26.1
Lung ca. SHP-77	61.6	CNS cancer (glio) SF-295	34.2
Lung ca. A549	39.2	Brain (Amygdala) Pool	2.6
Lung ca. NCI-H526	13.4	Brain (cerebellum)	3.5
Lung ca. NCI-H23	57.8	Brain (fetal)	3.1
Lung ca. NCI-H460	12.2	Brain (Hippocampus) Pool	3.7
Lung ca. HOP-62	8.8	Cerebral Cortex Pool	4.3
Lung ca. NCI-H522	39.8	Brain (Substantia nigra) Pool	2.8
Liver	1.3	Brain (Thalamus) Pool	4.5
Fetal Liver	6.4	Brain (whole)	1.8
Liver ca. HepG2	8.4	Spinal Cord Pool	3.4
Kidney Pool	15.7	Adrenal Gland	9.7
Fetal Kidney	11.3	Pituitary gland Pool	5.2
Renal ca. 786-0	50.7	Salivary Gland	11.1
Renal ca. A498	17.6	Thyroid (female)	10.0
Renal ca. ACHN	26.6	Pancreatic ca. CAPAN2	26.1
Renal ca. UO-31	44.1	Pancreas Pool	4.9

General_screening_panel_v1.6 Summary: Ag6911 Highest expression of this gene is seen in a breast cancer cell line (CT=29.8). This gene is widely expressed in this panel, with moderate expression seen in brain, colon, gastric, lung, breast, ovarian, and melanoma cancer cell lines. This expression profile suggests a role for this gene product in cell survival and proliferation. Modulation of this gene product may be useful in the treatment of cancer.

Among tissues with metabolic function, this gene is expressed at low but significant levels in pituitary, adipose, adrenal gland, pancreas, thyroid, fetal liver and adult and fetal skeletal muscle and heart. This widespread expression among these tissues suggests that this

gene product may play a role in normal neuroendocrine and metabolic function and that dysregulated expression of this gene may contribute to neuroendocrine disorders or metabolic diseases, such as obesity and diabetes.

This gene is also expressed at low but significant levels in the CNS, including the hippocampus, thalamus, substantia nigra, cerebellum and cerebral cortex. Therefore, therapeutic modulation of the expression or function of this gene may be useful in the treatment of neurologic disorders, such as Alzheimer's disease, Parkinson's disease, schizophrenia, multiple sclerosis, stroke and epilepsy.

AB. NOV91c and NOV91b: FIP-2 like

Expression of full-length physical clones NOV91c and variant NOV91b was assessed using the primer-probe set Ag6162, described in Table ABA.

Table ABA. Probe Name Ag6162

Primers	Sequences	Length	Start Position	SEQ ID No
Forward	5'-ttgtgtgcatctgtacacagta-3'	25	746	432
Probe	TET-5'-tggacttttcacctctgttttagcc-3'-TAMRA	26	774	433
Reverse	5'-gctatcagaaatcaaaatggaaca-3'	24	800	434

15

Example D: Identification of Single Nucleotide Polymorphisms in NOVX nucleic acid sequences

Variant sequences are also included in this application. A variant sequence can include a single nucleotide polymorphism (SNP). A SNP can, in some instances, be referred to as a "cSNP" to denote that the nucleotide sequence containing the SNP originates as a cDNA. A SNP can arise in several ways. For example, a SNP may be due to a substitution of one nucleotide for another at the polymorphic site. Such a substitution can be either a transition or a transversion. A SNP can also arise from a deletion of a nucleotide or an insertion of a nucleotide, relative to a reference allele. In this case, the polymorphic site is a site at which one allele bears a gap with respect to a particular nucleotide in another allele. SNPs occurring within genes may result in an alteration of the amino acid encoded by the gene at the position of the SNP. Intragenic SNPs may also be silent, when a codon including a SNP encodes the same amino acid as a result of the redundancy of the genetic code. SNPs occurring outside the region of a gene, or in an intron within a gene, do not result in changes in any amino acid sequence of a protein but may result in altered regulation of the expression

pattern. Examples include alteration in temporal expression, physiological response regulation, cell type expression regulation, intensity of expression, and stability of transcribed message.

SeqCalling assemblies produced by the exon linking process were selected and
5 extended using the following criteria. Genomic clones having regions with 98% identity to all or part of the initial or extended sequence were identified by BLASTN searches using the relevant sequence to query human genomic databases. The genomic clones that resulted were selected for further analysis because this identity indicates that these clones contain the genomic locus for these SeqCalling assemblies. These sequences were analyzed for putative
10 coding regions as well as for similarity to the known DNA and protein sequences. Programs used for these analyses include Grail, Genscan, BLAST, HMMER, FASTA, Hybrid and other relevant programs.

Some additional genomic regions may have also been identified because selected SeqCalling assemblies map to those regions. Such SeqCalling sequences may have
15 overlapped with regions defined by homology or exon prediction. They may also be included because the location of the fragment was in the vicinity of genomic regions identified by similarity or exon prediction that had been included in the original predicted sequence. The sequence so identified was manually assembled and then may have been extended using one or more additional sequences taken from CuraGen Corporation's human SeqCalling database.
20 SeqCalling fragments suitable for inclusion were identified by the CuraTools™ program SeqExtend or by identifying SeqCalling fragments mapping to the appropriate regions of the genomic clones analyzed.

The regions defined by the procedures described above were then manually integrated and corrected for apparent inconsistencies that may have arisen, for example, from miscalled
25 bases in the original fragments or from discrepancies between predicted exon junctions, EST locations and regions of sequence similarity, to derive the final sequence disclosed herein. When necessary, the process to identify and analyze SeqCalling assemblies and genomic clones was reiterated to derive the full length sequence (Alderborn et al., Determination of Single Nucleotide Polymorphisms by Real-time Pyrophosphate DNA Sequencing. Genome
30 Research. 10 (8) 1249-1265, 2000).

Variants are reported individually but any combination of all or a select subset of variants are also included as contemplated NOVX embodiments of the invention.

NOV1a SNP data:

NOV1a has 4 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 1 and 2, respectively. The nucleotide sequence of the NOV1a variants differ as shown in Table DA.

5

Table DA. cSNP and Coding Variants for NOV1a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380191	87	G	A	6	Ser	Asn
13380274	155	T	C	29	Ser	Pro
13380190	186	C	T	39	Ser	Leu
13380275	200	C	T	44	His	Tyr

NOV2a SNP data:

NOV2a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 3 and 4, respectively. The nucleotide sequence of the NOV2a variants differ as shown in Table DB.

10

Table DB. cSNP and Coding Variants for NOV2a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380382	1046	G	T	307	Val	Val

15

NOV6a SNP data:

NOV6a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 19 and 20, respectively. The nucleotide sequence of the NOV6a variants differ as shown in Table DC.

Table DC. cSNP and Coding Variants for NOV6a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380329	73	C	T	17	Pro	Ser

13380317	180	C	T	52	Cys	Cys
13380318	323	T	C	100	Val	Ala

NOV7a SNP data:

NOV7a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs:23 and 24, respectively. The nucleotide sequence of the NOV7a variants differ as shown in Table DD.

Table DD. cSNP and Coding Variants for NOV7a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380328	881	A	G	283	Thr	Ala
13380327	1429	G	A	465	Glu	Glu

10 **NOV8a SNP data:**

NOV8a has 8 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 33 and 34, respectively. The nucleotide sequence of the NOV8a variants differ as shown in Table DE.

Table DE. cSNP and Coding Variants for NOV8a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380302	360	T	C	8	Asp	Asp
13380301	545	A	G	70	Asn	Ser
13380300	677	G	A	114	Arg	Lys
13380299	1433	A	C	0		
13380298	1470	C	T	0		
13380297	1711	G	C	0		
13380296	1717	A	G	0		
13380295	1925	C	T	0		

NOV9b SNP data:

NOB9b has 7 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 37 and 38, respectively. The nucleotide sequence of the NOV9b variants differ as shown in Table DF.

5

Table DF. cSNP and Coding Variants for NOV9b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380197	47	C	T	16	Pro	Leu
13380196	258	G	A	86	Glu	Glu
13375934	309	C	T	103	Ala	Ala
13375935	340	G	A	114	Ala	Thr
13375936	393	T	C	131	Asp	Asp
13380199	435	G	A	145	Leu	Leu
13375938	457	G	A	153	Gly	Ser

NOV10a SNP data:

NOV10a has 24 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 39 and 40, respectively. The nucleotide sequence of the NOV10a variants differ as shown in Table DG.

10

Table DG. cSNP and Coding Variants for NOV10a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380337	63	G	C	6	Arg	Pro
13380342	220	C	T	58	Asn	Asn
13380343	282	G	A	79	Ser	Asn
13380344	299	C	T	85	Arg	Trp
13380345	303	G	A	86	Arg	Lys
13380346	345	C	T	100	Pro	Leu
13380347	362	A	G	106	Ser	Gly
13380348	445	T	C	133	Ile	Ile
13380349	520	G	A	158	Glu	Glu

13380350	554	G	T	170	Asp	Tyr
13380351	559	C	T	171	Thr	Thr
13380352	576	A	G	177	Gln	Arg
13380353	594	A	G	183	Glu	Gly
13380354	670	C	T	208	Ala	Ala
13380355	737	A	G	231	Ile	Val
13380356	772	C	T	242	Asn	Asn
13380357	773	C	T	243	Arg	End
13380358	886	T	C	280	Arg	Arg
13380307	1120	C	T	358	Thr	Thr
13380168	1157	C	T	371	Leu	Phe
13380167	1265	C	T	407	Leu	Leu
13380305	1273	T	C	409	His	His
13380306	1307	C	T	0		
13380359	1309	G	A	0		

5 **NOV11a SNP data:**

NOV11a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 43 and 44, respectively. The nucleotide sequence of the NOV11a variants differ as shown in Table DH.

Table DH. cSNP and Coding Variants for NOV11a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380247	57	T	C	0		
13380230	264	C	A	50	Thr	Thr
13380304	640	A	G	176	Met	Val

NOV12b SNP data:

NOV12b has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 47 and 48, respectively. The nucleotide sequence of the NOV12b variants differ as shown in Table DI.

5

Table DI. cSNP and Coding Variants for NOV12b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380285	366	C	T	0		

NOV13a SNP data:

NOV13a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 49 and 50, respectively. The nucleotide sequence of the NOV13a variants differ as shown in Table DJ.

10

Table DJ. cSNP and Coding Variants for NOV13a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380334	91	C	A	16	Ala	Glu
13380333	997	C	T	318	Ser	Phe

15

NOV14a SNP data:

NOV14a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 51 and 52, respectively. The nucleotide sequence of the NOV14a variants differ as shown in Table DK.

Table DK. cSNP and Coding Variants for NOV14a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380308	147	T	C	44	Val	Ala
13380309	597	C	T	194	Pro	Leu
13380310	786	T	C	257	Leu	Pro

20

NOV16a SNP data:

NOV16a has 5 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 57 and 58, respectively. The nucleotide sequence of the NOV16a variants differ as shown in Table DL.

Table DL. cSNP and Coding Variants for NOV16a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380246	575	A	G	176	Asn	Ser
13380245	603	C	T	185	Gly	Gly
13380239	1175	C	A	376	Pro	Gln
13380238	1253	G	A	402	Ser	Asn
13380237	1390	T	G	448	Leu	Val

NOV18a SNP data:

NOV18a has 5 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 61 and 62, respectively. The nucleotide sequence of the NOV18a variants differ as shown in Table DM.

Table DM. cSNP and Coding Variants for NOV18a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380324	42	T	G	0		
13380323	163	A	G	32	Thr	Ala
13380289	3363	A	G	1098	Glu	Glu
13380226	3489	C	T	1140	Ser	Ser
13380227	3782	C	A	1238	Thr	Lys

NOV19a SNP data:

NOV19a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 67 and 68, respectively. The nucleotide sequence of the NOV19a variants differ as shown in Table DN.

Table DN. cSNP and Coding Variants for NOV19a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380176	260	G	A	69	Ser	Ser
13380177	371	T	C	106	Ile	Ile

NOV20a SNP data:

- 5 NOV20a has 4 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 71 and 72, respectively. The nucleotide sequence of the NOV20a variants differ as shown in Table DO.

Table DO. cSNP and Coding Variants for NOV20a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380380	147	A	G	34	Pro	Pro
13380381	379	C	T	112	Pro	Ser
13380325	2939	A	C	965	His	Pro
13380331	4433	C	T	1463	Ala	Val

10

NOV21a SNP data:

NOV21a has 9 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 85 and 86, respectively. The nucleotide sequence of the NOV21a variants differ as shown in Table DP.

15

Table DP. cSNP and Coding Variants for NOV21a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380404	75	C	A	23	Pro	Thr
13380403	155	C	A	49	Pro	Pro

13380402	289	A	G	94	Asn	Ser
13380401	292	A	G	95	Asp	Gly
13380400	331	C	T	108	Ser	Leu
13380399	363	T	C	119	Tyr	His
13380398	370	G	A	121	Trp	End
13380397	536	G	C	176	Val	Val
13380396	548	C	G	180	Tyr	End

NOV22a SNP data:

- NOV22a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 89 and 90, respectively. The nucleotide sequence of the NOV22a variants differ as shown in Table DQ.

Table DQ. cSNP and Coding Variants for NOV22a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380210	172	T	A	48	Cys	Ser
13380209	342	A	C	104	Ser	Ser

10 **NOV23a SNP data:**

- NOV23a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 91 and 92, respectively. The nucleotide sequence of the NOV23a variants differ as shown in Table DR.

Table DR. cSNP and Coding Variants for NOV23a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380218	647	T	C	143	Thr	Thr
13380219	730	C	T	171	Thr	Ile

15

NOV24a SNP data:

NOV24a has 4 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 95 and 96, respectively. The nucleotide sequence of the NOV24a variants differ as shown in Table DS.

5

Table DS. cSNP and Coding Variants for NOV24a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380312	742	G	A	242	Glu	Lys
13380313	960	C	G	314	Thr	Thr
13380314	1144	G	C	0		
13380315	1462	G	C	0		

NOV25a SNP data:

NOV25a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 97 and 98, respectively. The nucleotide sequence of the NOV25a variants differ as shown in Table DT.

10

Table DT. cSNP and Coding Variants for NOV25a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380290	1958	A	G	568	Ala	Ala
13380291	2017	A	C	588	Asp	Ala
13380292	2094	C	T	0		

NOV27a SNP data:

NOV27a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 107 and 108, respectively. The nucleotide sequence of the NOV27a variants differ as shown in Table DU.

15

Table DU. cSNP and Coding Variants for NOV27a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380387	112	A	G	0		

13380385	948	A	G	273	Pro	Pro
----------	-----	---	---	-----	-----	-----

NOV28a SNP data:

- NOV28a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 111 and 112, respectively. The nucleotide sequence of the NOV28a variants differ as shown in Table DV.

Table DV. cSNP and Coding Variants for NOV28a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380287	265	A	C	25	Ala	Ala
13380288	736	G	T	0		

10 **NOV29a SNP data:**

NOV29a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 123 and 124, respectively. The nucleotide sequence of the NOV29a variants differ as shown in Table DW.

Table DW. cSNP and Coding Variants for NOV29a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377405	1685	G	A	555	Arg	Gln
13380222	2304	T	C	0		
13380223	2432	G	A	0		

15

NOV31a SNP data:

- NOV31a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 129 and 130, respectively. The nucleotide sequence of the NOV31a variants differ as shown in Table DX.

Table DX. cSNP and Coding Variants for NOV31a						
Variant	Nucleotides			Amino Acids		

	Position	Initial	Modified	Position	Initial	Modified
13380276	441	C	T	0		

NOV32b SNP data:

- NOV32b has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 133 and 134, respectively. The nucleotide sequence of the NOV32b variants differ as shown in Table DY.

Table DY. cSNP and Coding Variants for NOV32b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380407	1810	G	A	311	Ala	Thr
13375612	2273	C	T	465	Thr	Met

10 **NOV34a SNP data:**

NOV34a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 137 and 138, respectively. The nucleotide sequence of the NOV34a variants differ as shown in Table DZ.

Table DZ. cSNP and Coding Variants for NOV34a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380213	635	G	T	205	Cys	Phe
13380211	693	T	C	224	His	His

15

NOV36b SNP data:

- NOV36b has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 147 and 148, respectively. The nucleotide sequence of the NOV36b variants differ as shown in Table DAA.

Table DAA. cSNP and Coding Variants for NOV36b						
Variant	Nucleotides			Amino Acids		

	Position	Initial	Modified	Position	Initial	Modified
13380589	701	A	G	229	Gln	Arg
13380590	1413	T	C	466	Gly	Gly
13380591	2205	G	A	730	Gln	Gln

NOV37b SNP data:

- NOV37b has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 151 and 152, respectively. The nucleotide sequence of the NOV37b variants differ as shown in Table DAB.

Table DAB. cSNP and Coding Variants for NOV37b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380597	157	G	T	0	..	
13380626	1440	T	C	416	Ile	Thr

NOV38a SNP data:

- NOV38a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 155 and 156, respectively. The nucleotide sequence of the NOV38a variants differ as shown in Table DAC.

Table DAC. cSNP and Coding Variants for NOV38a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380602	201	A	G	56	Thr	Thr
13380596	285	A	T	84	Ser	Ser

NOV41a SNP data:

- NOV41a has 4 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 161 and 162, respectively. The nucleotide sequence of the NOV41a variants differ as shown in Table DAD.

Table DAD. cSNP and Coding Variants for NOV41a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380621	1499	A	C	494	Thr	Thr
13380620	1605	G	A	530	Glu	Lys
13380619	1649	T	C	544	Val	Val
13380618	2071	T	C	0		

NOV42a SNP data:

- NOV42a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 163 and 164, respectively. The nucleotide sequence of the NOV42a variants differ as shown in Table DAE.

Table DAE. cSNP and Coding Variants for NOV42a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380613	2607	G	A	841	Ala	Thr

NOV43a SNP data:

- NOV43a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 167 and 168, respectively. The nucleotide sequence of the NOV43a variants differ as shown in Table DAF.

Table DAF. cSNP and Coding Variants for NOV43a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380595	261	C	T	82	Tyr	Tyr
13380625	2989	T	C	992	Tyr	His

NOV44c SNP data:

- NOV44c has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 173 and 174, respectively. The nucleotide sequence of the NOV44c variants differ as shown in Table DAG.

Table DAG. cSNP and Coding Variants for NOV44c						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380603	246	G	A	73	Gly	Ser

NOV45a SNP data:

NOV45a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 175 and 176, respectively. The nucleotide sequence of the NOV45a variants differ as shown in Table DAH.

Table DAH. cSNP and Coding Variants for NOV45a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380607	1307	G	T	436	Ser	Ile

NOV46a SNP data:

NOV46a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 177 and 178, respectively. The nucleotide sequence of the NOV46a variants differ as shown in Table DAI.

Table DAI. cSNP and Coding Variants for NOV46a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380612	815	T	C	231	Leu	Ser
13380611	1028	C	T	302	Pro	Leu
13380609	1815	T	G	0		

NOV47a SNP data:

NOV47a has 9 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 179 and 180, respectively. The nucleotide sequence of the NOV47a variants differ as shown in Table DAJ.

Table DAJ. cSNP and Coding Variants for NOV47a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified

13380423	271	A	G	69	Lys	Glu
13380419	1869	G	A	601	Met	Ile
13380418	1921	A	G	619	Arg	Gly
13380417	1939	G	A	625	Asp	Asn
13380416	1960	C	G	632	Pro	Ala
13380414	2094	C	T	676	Pro	Pro
13380411	2146	G	A	694	Gly	Arg
13380409	2317	A	G	751	Thr	Ala
13380408	2615	G	T	0		

NOV48b SNP data:

NOV48b has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 183 and 184, respectively. The

5 nucleotide sequence of the NOV48b variants differ as shown in Table DAK.

Table DAK. cSNP and Coding Variants for NOV48b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377406	274	A	G	23	Ile	Val
13380605	294	G	C	29	Arg	Arg
13377408	821	A	G	205	Glu	Gly

NOV49a SNP data:

NOV49a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 185 and 186, respectively. The

10 nucleotide sequence of the NOV49a variants differ as shown in Table DAL.

Table DAL. cSNP and Coding Variants for NOV49a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380598	806	T	C	269	Val	Ala
13380599	878	G	A	293	Arg	His

NOV50a SNP data:

NOV50a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 189 and 190, respectively. The nucleotide sequence of the NOV50a variants differ as shown in Table DAM.

5

Table DAM. cSNP and Coding Variants for NOV50a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380616	85	A	G	17	Asp	Gly
13380614	576	C	T	181	Pro	Ser

NOV54a SNP data:

NOV54a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 203 and 204, respectively. The nucleotide sequence of the NOV54a variants differ as shown in Table DAN.

10

Table DAN. cSNP and Coding Variants for NOV54a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380622	1971	G	C	657	Ala	Ala
13380594	3003	C	T	1001	Asp	Asp

15

NOV52a SNP data:

NOV52a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 197 and 198, respectively. The nucleotide sequence of the NOV52a variants differ as shown in Table DAO.

Table DAO. cSNP and Coding Variants for NOV52a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380627	561	C	T	183	Ala	Val

20

NOV61a SNP data:

NOV61a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 227 and 228, respectively. The nucleotide sequence of the NOV61a variants differ as shown in Table DAP.

5

Table DAP. cSNP and Coding Variants for NOV61a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13379675	261	T	A	85	Phe	Ile

NOV62a SNP data:

NOV62a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 229 and 230, respectively. The nucleotide sequence of the NOV62a variants differ as shown in Table DAQ.

10

Table DAQ. cSNP and Coding Variants for NOV62a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380689	236	T	C	55	Cys	Arg
13380630	340	T	C	89	Asp	Asp

15

NOV66a SNP data:

NOV66a has 2 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 241 and 242, respectively. The nucleotide sequence of the NOV66a variants differ as shown in Table DAR.

Table DAR. cSNP and Coding Variants for NOV66a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380657	1659	A	G	551	Gln	Arg
13380658	1753	G	A	582	Leu	Leu

20

NOV67a SNP data:

NOV67a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 245 and 246, respectively. The nucleotide sequence of the NOV67a variants differ as shown in Table DAS.

5

Table DAS. cSNP and Coding Variants for NOV67a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380649	1293	C	A	427	Ala	Ala

NOV71a SNP data:

NOV71a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 253 and 254, respectively. The nucleotide sequence of the NOV71a variants differ as shown in Table DAT.

10

Table DAT. cSNP and Coding Variants for NOV71a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380681	4458	T	C	1484	Arg	Arg

NOV72a SNP data:

NOV72a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 255 and 256, respectively. The nucleotide sequence of the NOV72a variants differ as shown in Table DAU.

15

Table DAU. cSNP and Coding Variants for NOV72a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380650	3106	C	T	1005	Leu	Leu

NOV74a SNP data:

NOV74a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 265 and 266, respectively. The nucleotide sequence of the NOV74a variants differ as shown in Table DAV.

20

Table DAV. cSNP and Coding Variants for NOV74a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380635	5439	A	G	1361	Glu	Gly
13380634	5825	A	G	1490	Thr	Ala
13380633	11081	G	C	0		

NOV77a SNP data:

- NOV77a has 4 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 271 and 272, respectively. The nucleotide sequence of the NOV77a variants differ as shown in Table DAW.

Table DAW. cSNP and Coding Variants for NOV77a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380655	414	G	A	31	Val	Ile
13380654	955	A	G	211	Gln	Arg
13380653	971	C	T	216	Asp	Asp
13380643	1800	A	G	493	Ser	Gly

NOV80a SNP data:

NOV80a has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 277 and 278, respectively. The nucleotide sequence of the NOV80a variants differ as shown in Table DAX.

Table DAX. cSNP and Coding Variants for NOV80a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380632	940	T	G	0		

NOV81b SNP data:

NOV81b has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 281 and 282, respectively. The nucleotide sequence of the NOV81b variants differ as shown in Table DAY.

5

Table DAY. cSNP and Coding Variants for NOV81b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380667	385	C	T	9	Arg	Trp

NOV82b SNP data:

NOV82b has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 285 and 286, respectively. The nucleotide sequence of the NOV82b variants differ as shown in Table DAZ.

10

Table DAZ. cSNP and Coding Variants for NOV82b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380679	218	G	A	67	Ala	Thr
13380641	263	A	G	82	Lys	Glu
13380644	307	T	C	96	Leu	Leu

15

NOV83b SNP data:

NOV83b has 1 SNP variant, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 291 and 292, respectively. The nucleotide sequence of the NOV83b variants differ as shown in Table DBA.

Table DBA. cSNP and Coding Variants for NOV83b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377534	73	C	T	0		

20

NOV84a SNP data:

NOV84a has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 293 and 294, respectively. The nucleotide sequence of the NOV84a variants differ as shown in Table DBB.

5

Table DBB. cSNP and Coding Variants for NOV84a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13380665	1934	G	A	632	Asp	Asn
13380663	2176	T	C	712	Ser	Ser
13380662	2544	C	T	835	Ala	Val

NOV86b SNP data:

NOV86b has 6 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 305 and 306, respectively. The nucleotide sequence of the NOV86b variants differ as shown in Table DBC.

10

Table DBC. cSNP and Coding Variants for NOV86b						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13377818	97	A	G	33	Ser	Gly
13377817	867	A	C	289	Gln	His
13377816	909	A	G	303	Ile	Met
13377815	919	C	T	307	Pro	Ser
13377814	926	T	A	309	Leu	His
13377813	1122	T	C	374	Ser	Ser

15

NOV87a SNP data:

NOV87a has 6 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 307 and 308, respectively. The nucleotide sequence of the NOV87a variants differ as shown in Table DBD.

Table DBD. cSNP and Coding Variants for NOV87a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13375628	421	T	C	51	Ser	Pro
13375629	557	G	A	96	Gly	Asp
13375630	562	T	C	98	Ser	Pro
13380645	916	A	G	0		
13380646	1079	A	G	0		
13380647	1134	A	G	0		

NOV88a SNP data:

NOV88a has 5 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOa: 317 and 318, respectively. The nucleotide sequence of the NOV88a variants differ as shown in Table DBE.

Table DBE. cSNP and Coding Variants for NOV88a						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified
13379305	583	G	A	144	Ala	Thr
13379294	938	A	G	262	Gln	Arg
13379295	1195	A	G	348	Ile	Val
13379304	1367	T	C	405	Leu	Pro
13379296	1757	A	G	535	Tyr	Cys

NOV91c SNP data:

NOV91c has 3 SNP variants, whose variant positions for its nucleotide and amino acid sequences is numbered according to SEQ ID NOs: 333 and 334, respectively. The nucleotide sequence of the NOV91c variants differ as shown in Table DBF.

Table DBF. cSNP and Coding Variants for NOV91c						
Variant	Nucleotides			Amino Acids		
	Position	Initial	Modified	Position	Initial	Modified

13378928	278	A	G	67	Glu	Gly
13380669	1158	A	G	360	Glu	Glu
13378927	1671	G	A	0		

OTHER EMBODIMENTS

5 Although particular embodiments have been disclosed herein in detail, this has been done by way of example for purposes of illustration only, and is not intended to be limiting with respect to the scope of the appended claims, which follow. In particular, it is contemplated by the inventors that various substitutions, alterations, and modifications may be made to the invention without departing from the spirit and scope of the invention as

10 defined by the claims. The choice of nucleic acid starting material, clone of interest, or library type is believed to be a matter of routine for a person of ordinary skill in the art with knowledge of the embodiments described herein. Other aspects, advantages, and modifications considered to be within the scope of the following claims. The claims presented are representative of the inventions disclosed herein. Other, unclaimed inventions

15 are also contemplated. Applicants reserve the right to pursue such inventions in later claims.

CLAIMS

What is claimed is:

1. An isolated polypeptide comprising the mature form of an amino acid sequenced selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172.
2. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172.
3. An isolated polypeptide comprising an amino acid sequence which is at least 95% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172.
4. An isolated polypeptide, wherein the polypeptide comprises an amino acid sequence comprising one or more conservative substitutions in the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172.
5. The polypeptide of claim 1 wherein said polypeptide is naturally occurring.
6. A composition comprising the polypeptide of claim 1 and a carrier.
7. A kit comprising, in one or more containers, the composition of claim 6.
8. The use of a therapeutic in the manufacture of a medicament for treating a syndrome associated with a human disease, the disease selected from a pathology associated with the polypeptide of claim 1, wherein the therapeutic comprises the polypeptide of claim 1.
9. A method for determining the presence or amount of the polypeptide of claim 1 in a sample, the method comprising:

- (a) providing said sample;
- (b) introducing said sample to an antibody that binds immunospecifically to the polypeptide; and
- (c) determining the presence or amount of antibody bound to said polypeptide, thereby determining the presence or amount of polypeptide in said sample.

10. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the polypeptide of claim 1 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the polypeptide in a sample from the first mammalian subject; and
- b) comparing the expression of said polypeptide in the sample of step (a) to the expression of the polypeptide present in a control sample from a second mammalian subject known not to have, or not to be predisposed to, said disease,

wherein an alteration in the level of expression of the polypeptide in the first subject as compared to the control sample indicates the presence of or predisposition to said disease.

11. A method of identifying an agent that binds to the polypeptide of claim 1, the method comprising:

- (a) introducing said polypeptide to said agent; and
- (b) determining whether said agent binds to said polypeptide.

12. The method of claim 11 wherein the agent is a cellular receptor or a downstream effector.

13. A method for identifying a potential therapeutic agent for use in treatment of a pathology, wherein the pathology is related to aberrant expression or aberrant physiological interactions of the polypeptide of claim 1, the method comprising:

- (a) providing a cell expressing the polypeptide of claim 1 and having a property or function ascribable to the polypeptide;
- (b) contacting the cell with a composition comprising a candidate substance; and
- (c) determining whether the substance alters the property or function ascribable to the polypeptide;

whereby, if an alteration observed in the presence of the substance is not observed when the cell is contacted with a composition in the absence of the substance, the substance is identified as a potential therapeutic agent.

14. A method for screening for a modulator of activity of or of latency or predisposition to a pathology associated with the polypeptide of claim 1, said method comprising:

- (a) administering a test compound to a test animal at increased risk for a pathology associated with the polypeptide of claim 1, wherein said test animal recombinantly expresses the polypeptide of claim 1;
- (b) measuring the activity of said polypeptide in said test animal after administering the compound of step (a); and
- (c) comparing the activity of said polypeptide in said test animal with the activity of said polypeptide in a control animal not administered said polypeptide, wherein a change in the activity of said polypeptide in said test animal relative to said control animal indicates the test compound is a modulator activity of or latency or predisposition to, a pathology associated with the polypeptide of claim 1.

15. The method of claim 14, wherein said test animal is a recombinant test animal that expresses a test protein transgene or expresses said transgene under the control of a promoter at an increased level relative to a wild-type test animal, and wherein said promoter is not the native gene promoter of said transgene.

16. A method for modulating the activity of the polypeptide of claim 1, the method comprising contacting a cell sample expressing the polypeptide of claim 1 with a compound that binds to said polypeptide in an amount sufficient to modulate the activity of the polypeptide.

17. A method of treating or preventing a pathology associated with the polypeptide of claim 1, the method comprising administering the polypeptide of claim 1 to a subject in which such treatment or prevention is desired in an amount sufficient to treat or prevent the pathology in the subject.

18. The method of claim 17, wherein the subject is a human.
19. A method of treating a pathological state in a mammal, the method comprising administering to the mammal a polypeptide in an amount that is sufficient to alleviate the pathological state, wherein the polypeptide is a polypeptide having an amino acid sequence at least 95% identical to a polypeptide comprising the amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172 or a biologically active fragment thereof.
20. An isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 172.
21. The nucleic acid molecule of claim 20, wherein the nucleic acid molecule is naturally occurring.
22. A nucleic acid molecule, wherein the nucleic acid molecule differs by a single nucleotide from a nucleic acid sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 172.
23. An isolated nucleic acid molecule encoding the mature form of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:2n, wherein n is an integer between 1 and 172.
24. An isolated nucleic acid molecule comprising a nucleic acid selected from the group consisting of 2n-1, wherein n is an integer between 1 and 172.
25. The nucleic acid molecule of claim 20, wherein said nucleic acid molecule hybridizes under stringent conditions to the nucleotide sequence selected from the group consisting of SEQ ID NO: 2n-1, wherein n is an integer between 1 and 172, or a complement of said nucleotide sequence.
26. A vector comprising the nucleic acid molecule of claim 20.

27. The vector of claim 26, further comprising a promoter operably linked to said nucleic acid molecule.

28. A cell comprising the vector of claim 26.

29. An antibody that immunospecifically binds to the polypeptide of claim 1.

30. The antibody of claim 29, wherein the antibody is a monoclonal antibody.

31. The antibody of claim 29, wherein the antibody is a humanized antibody.

32. A method for determining the presence or amount of the nucleic acid molecule of claim 20 in a sample, the method comprising:

- (a) providing said sample;
- (b) introducing said sample to a probe that binds to said nucleic acid molecule;
and
- (c) determining the presence or amount of said probe bound to said nucleic acid molecule,

thereby determining the presence or amount of the nucleic acid molecule in said sample.

33. The method of claim 32 wherein presence or amount of the nucleic acid molecule is used as a marker for cell or tissue type.

34. The method of claim 33 wherein the cell or tissue type is cancerous.

35. A method for determining the presence of or predisposition to a disease associated with altered levels of expression of the nucleic acid molecule of claim 20 in a first mammalian subject, the method comprising:

- a) measuring the level of expression of the nucleic acid in a sample from the first mammalian subject; and
- b) comparing the level of expression of said nucleic acid in the sample of step (a) to the level of expression of the nucleic acid present in a control sample from a second mammalian subject known not to have or not be predisposed to, the disease;

wherein an alteration in the level of expression of the nucleic acid in the first subject as compared to the control sample indicates the presence of or predisposition to the disease.

36. A method of producing the polypeptide of claim 1, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 172.

37. The method of claim 36 wherein the cell is a bacterial cell.

38. The method of claim 36 wherein the cell is an insect cell.

39. The method of claim 36 wherein the cell is a yeast cell.

40. The method of claim 36 wherein the cell is a mammalian cell.

41. A method of producing the polypeptide of claim 2, the method comprising culturing a cell under conditions that lead to expression of the polypeptide, wherein said cell comprises a vector comprising an isolated nucleic acid molecule comprising a nucleic acid sequence selected from the group consisting of SEQ ID NO:2n-1, wherein n is an integer between 1 and 172.

42. The method of claim 41 wherein the cell is a bacterial cell.

43. The method of claim 41 wherein the cell is an insect cell.

44. The method of claim 41 wherein the cell is a yeast cell.

45. The method of claim 41 wherein the cell is a mammalian cell.

**This Page is Inserted by IFW Indexing and Scanning
Operations and is not part of the Official Record**

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images include but are not limited to the items checked:

- ☐ BLACK BORDERS
- ☐ IMAGE CUT OFF AT TOP, BOTTOM OR SIDES
- ☐ FADED TEXT OR DRAWING
- ☒ BLURRED OR ILLEGIBLE TEXT OR DRAWING
- ☐ SKEWED/SLANTED IMAGES
- ☐ COLOR OR BLACK AND WHITE PHOTOGRAPHS
- ☐ GRAY SCALE DOCUMENTS
- ☐ LINES OR MARKS ON ORIGINAL DOCUMENT
- ☐ REFERENCE(S) OR EXHIBIT(S) SUBMITTED ARE POOR QUALITY
- ☐ OTHER: _____

IMAGES ARE BEST AVAILABLE COPY.

As rescanning these documents will not correct the image problems checked, please do not report these problems to the IFW Image Problem Mailbox.